



# Estimation des flux d'immigration : réconciliation de deux sources par une approche bayésienne

Julyan Arbel, Vianney Costemalle

## ► To cite this version:

Julyan Arbel, Vianney Costemalle. Estimation des flux d'immigration : réconciliation de deux sources par une approche bayésienne. *Economie et Statistique / Economics and Statistics*, 2016, 483-484-485, pp.121-149. hal-01396606

**HAL Id: hal-01396606**

**<https://hal.science/hal-01396606>**

Submitted on 21 Dec 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Estimation des flux d'immigration: réconciliation de deux sources par une approche bayésienne

Julyan Arbel \* et Vianney Costemalle \*\*

Deux sources au moins permettent d'appréhender les flux migratoires en France. D'une part les titres de séjour délivrés aux migrants majeurs de nationalité étrangère fournissent les dates d'arrivées et les durées de titre de séjour des migrants en situation légale. D'autre part les enquêtes annuelles de *Recensement de la population* qui ont lieu chaque année depuis 2004 donnent une photographie de la population vivant en France au début de l'année, permettant en particulier de recueillir la date d'arrivée en France des personnes nées à l'étranger. En se restreignant à un champ commun aux deux sources, constitué des personnes de nationalité étrangère (hors Espace économique européen et Suisse), nées à l'étranger, arrivant en France métropolitaine âgées de 19 ans ou plus, l'article propose un modèle de réconciliation des données en estimant par inférence bayésienne des taux de présence à l'*Enquête annuelle de recensement* et des taux d'obtention de titre de séjour selon le nombre d'années passées sur le territoire. Après avoir identifié parmi les migrants qui obtiennent un titre de séjour ceux qui s'installent au moins un an en France, qui est la condition pour être recensé selon les définitions internationales, les estimations indiquent que tous les migrants entrant une année ne sont pas présents (ou identifiés) au recensement la première année et qu'à partir de la troisième année, la part de ceux qui sont présents au recensement diminue en moyenne de 3 % par an (notamment du fait des sorties du territoire). 70 % de ces nouveaux migrants obtiennent leur titre de séjour l'année de leur arrivée ou l'année suivante. De fortes disparités existent entre nationalités quant à la présence estimée au recensement. De même, les femmes semblent plus souvent recensées que les hommes, et les plus jeunes que les plus âgés. Enfin, l'analyse fine des deux sources permet de comprendre leurs forces et faiblesses respectives en vue d'estimer les flux migratoires.

## Rappel :

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux mêmes, et non les institutions auxquelles ils appartiennent, ni a fortiori l'Insee.

Codes JEL : C11, C13, J11.

Mots clés : immigration, inférence bayésienne, titres de séjour, recensement de la population.

\* Collegio Carlo Alberto, Turin, Université Bocconi, Milan et Centre de recherche en économie et statistique (Crest).

\*\* Insee.

Les auteurs remercient les membres de l'unité des études démographiques et sociales de l'Insee qui ont fait bénéficier ce travail de leurs avis judicieux, notamment sous l'angle du développement de la méthodologie et de l'amélioration de la rédaction. Ils sont également particulièrement redevables à Hippolyte d'Albis, Didier Blanchet, Gérard Bouvier et Christian Robert. Ils adressent par ailleurs leurs remerciements aux membres du département des statistiques, des études et de la documentation du Ministère de l'intérieur (DSED) qui ont fourni les données relatives aux titres de séjour ainsi que de nombreuses explications. Ils ont enfin bénéficié des conseils avisés de deux rapporteurs de la revue.

Depuis les années 1990, les migrations internationales s'intensifient à travers le monde (Nations Unies, 2013) contribuant de plus en plus à modeler les évolutions démographiques des pays développés dont le solde naturel tend à devenir nul voire négatif<sup>1</sup>. Cette évolution de la migration avait été soulignée dès 1994 lors de la conférence mondiale des Nations Unies sur la population<sup>2</sup>. Il est donc important de disposer de mesures fiables des flux migratoires si l'on souhaite analyser ces phénomènes. Au niveau de l'Union européenne, il est difficile de fournir des statistiques sur les flux migratoires cohérentes d'un pays à l'autre, faute d'harmonisation des définitions et des méthodes employées (Poulain et Herm, 2002). Tous les pays membres de l'Union européenne, à l'exception de la France, du Royaume-Uni, de l'Irlande, du Portugal et de la Grèce, disposent de registres de la population, permettant en théorie d'enregistrer les arrivées comme les départs d'habitants et ainsi de donner une estimation précise des migrations (Poulain et Herm, 2002). En réalité la qualité de ces informations n'est pas toujours bonne, ce qui peut s'expliquer entre autre par le fait que les immigrants ont plus d'incitations à s'inscrire sur un nouveau registre que les émigrants à se désinscrire de leur ancien registre (Willekens, 1994). Pour les pays sans registres, le moyen le plus utilisé pour déterminer les flux migratoires reste le *Recensement de la population*. Il permet en effet d'une part de déterminer les nouveaux migrants en posant des questions sur le lieu de résidence antérieure, et d'autre part de disposer des effectifs de population résidente à des dates fixées. Connaissant la natalité, la mortalité et les naturalisations, on peut alors retrouver le solde migratoire des étrangers entre ces dates.

Dans la perspective d'améliorer la qualité des statistiques et de les harmoniser, l'ONU a exposé en 1998 ses *Recommandations en matière de statistiques des migrations internationales* (ONU, 1999), faisant notamment la distinction entre deux types de migrants. Le migrant de court terme, d'une part, est « une personne qui se rend dans un pays autre que celui de sa résidence habituelle pour une période d'au moins trois mois mais de moins d'un an, à l'exception des cas où le voyage dans ce pays est effectué à des fins de loisirs, de vacances, de visites à des amis ou à de la famille, d'affaires, de traitement médical ou de pèlerinage religieux ». D'autre part, le migrant de long terme est « une personne qui se rend dans un pays autre que celui de sa résidence habituelle pour une période d'au moins un an, de façon que le pays de destination

devienne effectivement son nouveau pays de résidence habituelle ».

Se basant sur ces définitions, le récent règlement européen relatif aux statistiques communautaires sur la migration et la protection internationale (Règlement 862/2007) définit l'immigration comme étant celle des migrants de long terme : « L'immigration est l'établissement, par une personne, de sa résidence habituelle, pour une période atteignant ou supposée atteindre douze mois au moins, sur le territoire d'un État membre après avoir eu précédemment sa résidence habituelle dans un autre État membre ou dans un pays tiers ». C'est cette définition de l'immigration qui est retenue pour cet article.

En France deux sources principales sont habituellement utilisées pour appréhender les phénomènes migratoires : les titres de séjour et le *Recensement de la population*. La première est de nature administrative puisqu'il s'agit de documents délivrés aux étrangers tout au long de l'année afin qu'ils puissent séjourner légalement en France et y avoir une activité professionnelle, tandis que la deuxième est une photographie de la population à un moment donné réalisée par enquête sur une sous-population. Traditionnellement, les chercheurs de l'Ined s'intéressent à la migration de long terme (selon les recommandations des Nations Unies, et ce dès leur parution) en considérant les personnes recevant un premier titre de séjour ayant une durée de validité supérieure ou égale à un an (Thierry, 2001 ; Mazuy *et al.*, 2013). Des services statistiques ministériels comme le DSED (Menard et Papon, 2009), la Drees (Bègue, 2007) ou la Dares (Dares, 2011) exploitent également les titres de séjours tandis que l'Insee fournit des résultats à partir des données du recensement (Brutel, 2014).

Ces deux sources de nature différente ne sont pas directement adaptées à l'estimation des flux migratoires, mais offrent chacune un éclairage particulier sur les phénomènes de migrations. Comment alors est-il possible de les réconcilier afin d'obtenir un panorama plus cohérent ? C'est l'objet de cet article qui construit un modèle intégrant les deux sources permettant

1. Voir (Sardon, 2006) pour ce qui est de l'Europe.

2. « La nature et l'ampleur des déplacements de population étant en train de changer, il est impératif d'entreprendre des recherches sur les causes et les conséquences de la mobilité et des migrations, aussi bien internes qu'internationales. », paragraphe 12,25 du rapport de la conférence internationale sur la population et le développement (Nations Unies, 1994).

d'estimer par inférence bayésienne, les flux des migrants par année ainsi que des taux de présence au recensement et de délivrance de titres de séjours.

D'autres pays ont essayé de combiner plusieurs sources pour estimer les flux migratoires. Aux États-Unis par exemple, trois sources ont été mobilisées (le recensement décennal, l'enquête annuelle sur la population actuelle et les déclarations de revenus) pour estimer les migrations inter-états à l'aide de modèles linéaires généralisés (Willekens, 1994). Au Royaume-Uni, les changements d'adresses enregistrés dans le National Health Service Central Register ont été couplés avec les données du recensement pour mieux estimer la migration interne à l'aide d'un modèle d'interaction spatiale dans lequel les flux observés par les deux sources résultent d'une loi de Poisson (Raymer *et al.*, 2007). Les paramètres du modèle sont alors estimés par maximum de vraisemblance. Au niveau européen, le projet MIMOSA (Migration Modelling for Statistical Analyses) (Raymer et Abel, 2008 ; de Beer, 2010) a développé un modèle permettant d'estimer une matrice de flux d'immigration et d'émigration entre plusieurs pays européens tenant compte de l'information apportée par chacun des pays étudiés. En effet, un flux peut être considéré sous deux angles : à partir du pays dont partent les migrants, ou bien à partir du pays où ils arrivent. Ce modèle permet de réconcilier les données entre pays présentant une forte hétérogénéité du fait de méthodes d'estimation et de définitions de l'immigration différentes. Il permet de plus d'estimer des flux manquants.

En ce qui concerne l'approche bayésienne, plusieurs auteurs l'ont adoptée pour résoudre des problèmes de démographie. Pour les projections de populations, cette approche permet d'introduire des modèles probabilistes là où il y avait des modèles déterministes (Alkema *et al.*, 2014). Des informations *a priori* peuvent être mobilisées pour estimer des tables de contingence représentant des flux migratoires (Albert et Gupta, 1983). Plus récemment, Bijak a développé des méthodes reposant sur le paradigme bayésien pour prévoir les migrations internationales en Europe (Bijak, 2011). Un nouveau projet européen, faisant suite au projet MIMOSA repose également sur un modèle bayésien. Il s'agit du projet Imem (*Integrated Modelling of European Migration*) (Raymer *et al.*, 2011) qui estime des matrices de flux d'immigration et d'émigration à l'aide d'un modèle log-linéaire, dont les paramètres sont estimés par inférence

bayésienne. Ce modèle prend en compte les différentes définitions de la migration, notamment en terme de temps de séjour. Il utilise de plus des covariables macroéconomiques pour expliquer les flux.

À notre connaissance, aucune estimation des flux migratoires internationaux basée sur le rapprochement entre les titres de séjours délivrés aux étrangers et un recensement n'a été effectuée en utilisant une approche bayésienne. Le modèle présenté ici est différent des modèles cités précédemment, même si l'idée générale est souvent la même, à savoir que les flux observés par différentes sources sont des observations partielles et complémentaires des flux réels.

Dans un premier temps, on décrit les spécificités de chacune des deux sources (la base des titres de séjour et les enquêtes annuelles de recensement) afin de réduire l'étude à un champ commun. On décrit notamment comment se rapprocher du champ des migrants de long terme dans la base des titres de séjour. Ensuite, on propose un modèle de réconciliation entre les deux sources, faisant apparaître des taux de présence au recensement et des taux d'obtention de titre. Afin que le modèle proposé soit identifiable, il sera alors nécessaire d'ajouter une contrainte sur les paramètres. On spécifiera alors la valeur de la proportion de migrants obtenant leur premier titre de séjour dans les 10 ans après leur arrivée sur le territoire. Puis on présente les résultats des estimations obtenues par inférence bayésienne. Enfin, on discute des limites du modèle.

## Les titres de séjour

Lorsqu'une personne majeure de nationalité de pays tiers<sup>3</sup> séjourne en France trois mois ou plus, elle doit posséder un titre de séjour délivré par la préfecture de son département de résidence. La base issue de l'Application de gestion des dossiers des ressortissants étrangers en France (base AGDREF) est une source administrative utilisée depuis 1993 pour créer et gérer les titres de séjours en préfecture. Les préfectures gèrent chaque année près d'un million de demandes de titres qui sont des

3. Les pays tiers sont ici les pays hors Espace économique européen (Union européenne, Islande, Norvège et Liechtenstein) et hors Suisse. Le champ des pays tiers se rétrécit donc au fur et à mesure que l'Union européenne s'élargit. Par souci de simplicité on garde dans cet article le dernier champ défini par l'Union européenne avant l'entrée de la Croatie en 2013 : il s'agit de l'UE à 27 ayant intégré la Bulgarie et la Roumanie en 2007.

demandes de premier titre (création de dossier), des demandes de renouvellement de titre ou des demandes de duplicatas. À ces titres s'ajoutent des documents provisoires qui donnent le droit de rester en France pour de courtes durées en général. Il s'agit principalement d'autorisations provisoires de séjour, valables de 1 à 6 mois, de récépissés en attendant que la demande de titre soit confirmée ou de convocations. Ces documents provisoires ne sont pas des titres de séjour et n'ouvrent pas au droit de travailler (sauf cas particuliers).

Les titres de séjour accordés aux étrangers majeurs diffèrent principalement par leur motif et leur durée de validité. On peut les regrouper en cinq grands motifs : économique, familial, étudiant, divers et humanitaire. La durée de validité peut varier de 3 mois à 10 ans. Il existe aussi des titres de séjour pour résident permanent ayant une durée de validité indéterminée : ce sont les cartes de résidents permanents. Chacun de ces motifs (mis à part le motif « étudiant ») peut ensuite être détaillé en sous-motifs. Les conditions d'obtention de ces titres sont détaillées dans le code de l'entrée et du séjour des étrangers et du droit d'asile (Ceseda). Lorsque le migrant arrive en France il doit demander la création d'un titre auprès de sa préfecture (on parle alors de premier titre, quelle que soit la durée de validité de ce titre) qui peut ensuite être prolongé par plusieurs renouvellements successifs. Les titres ainsi obtenus par renouvellement ne sont pas toujours identiques aux précédents, car plusieurs caractéristiques peuvent changer comme la durée de validité du titre, le motif du titre ou l'adresse du migrant. Par construction statistique, le DSED considère que tout en gardant le même numéro d'étranger, une personne peut cumuler au cours de sa vie plusieurs premiers titres si l'absence entre la fin de validité d'un titre et la demande d'un nouveau titre est supérieure à un an. On considère alors que la nouvelle demande de titre correspond à une entrée du migrant qui s'était établi hors de France pendant un an ou plus et qu'il ne s'agit donc pas d'un renouvellement. Chaque année<sup>4</sup> il y a en moyenne 200 000 personnes qui obtiennent un premier titre. Parmi elles, en moyenne 2 100 mineurs, dont 500 de 16 ans et 1 600 de 17 ans, demandent un titre pour pouvoir être en situation de travailler. Autrement, les mineurs sont dispensés par la loi de titre pour rester sur le territoire français<sup>5</sup>. De plus, il arrive que certains ressortissants de l'Union européenne demandent des titres de séjours, même s'ils n'en ont pas besoin pour séjourner en France<sup>6</sup>.

Depuis 2009 ont été mis en place des visas spéciaux considérés comme des titres de séjours (cf. encadré 1), dont une partie seulement est intégrée dans la base AGDREF. On utilise dans la suite de cette étude, les données de la base AGDREF correspondant à des titres délivrés entre 2004 et 2012.

Il est important de remarquer que la base AGDREF fournit des informations sur une partie des migrants seulement. Il s'agit des migrants majeurs dont la nationalité est celle d'un pays tiers, venant séjourner en France pour plus de 3 mois.

## Le recensement : comment traiter la non-réponse partielle?

Le Recensement de la population (RP) a lieu chaque année en France depuis 2004 sous forme d'une *Enquête annuelle de recensement* (EAR). Elle vise à compter l'ensemble de la population vivant en France<sup>7</sup> ainsi qu'à acquérir certaines informations sur cette population. Elle n'est pas exhaustive puisqu'elle s'effectue chaque année sur une sous-partie de la population sondée de façon aléatoire, avant d'être extrapolé à l'ensemble de la population<sup>8</sup>.

Aux personnes nées à l'étranger on pose, en plus de la question sur le pays de naissance, la question « En quelle année êtes vous arrivé(e) en France ? » permettant de compter le nombre de personnes étrangères et nées à l'étranger entrant chaque année. Le flux de l'année N est déterminé par le nombre de personnes recensées

4. Chiffres calculés à partir de la base AGDREF, sur les premiers titres délivrés en France métropolitaine entre les années 2004 et 2012. Dans toute cette étude, lorsqu'on parle d'année d'obtention de titre ou d'année de délivrance de titre, on se réfère en fait à l'année de début de validité de celui-ci (la demande de titre ayant pu dans les faits être acceptée plus tard, éventuellement après la fin de l'année civile du début de validité).

5. Néanmoins, certains mineurs possèdent des documents leur permettant de voyager à l'étranger. De plus, il est possible de déterminer les mineurs étrangers qui sont venus en France par regroupement familial. Les mineurs ne seront pourtant pas pris en compte ici, du fait de leur non obligation à détenir un titre de séjour et donc de leur non exhaustivité dans la base AGDREF.

6. Jusqu'à récemment, les ressortissants de Bulgarie et de Roumanie avaient par exemple toujours besoin d'un titre de séjour pour pouvoir travailler en France.

7. En théorie, selon l'article 2 du règlement n°763/2008 du parlement européen et du conseil, seules les personnes résidant depuis plus d'un an en France ou qui y résideront plus d'un an doivent être recensées (Union européenne, 2008).

8. Les pondérations des EAR ne sont pas exactement les mêmes que celles du Recensement de la Population (RP) pour lequel certains traitements sont ajoutés.



en janvier de l'année N+1 et ayant déclaré être entré en France l'année N. L'année d'arrivée en France n'étant pas toujours renseignée, il est alors utile de s'appuyer sur les questions « Où habitez-vous au premier janvier cinq ans auparavant ? » pour les *Enquêtes annuelles de recensement* (EAR) jusqu'à 2010 et « Où habitez-vous au premier janvier de l'année précédente ? » pour les *Enquêtes annuelles de recensement* à partir de 2011, pour déduire si l'individu est un nouveau migrant ou non.

Pourtant, même avec cette information supplémentaire, la date d'arrivée en France ne peut être obtenue avec certitude. Par exemple, si une personne recensée en janvier 2010 a déclaré habiter à l'étranger cinq ans auparavant et n'a pas donné de date d'entrée sur le territoire, on peut seulement être sûr qu'elle est entrée entre 2005 et 2009 inclus. Si au contraire cette personne déclare habiter en France 5 ans auparavant, on peut uniquement en déduire que son année d'entrée est inférieure ou égale à 2004. De plus, certaines déclarations d'année d'arrivée paraissent incohérentes avec la réponse donnée sur la résidence antérieure, comme lorsqu'un individu déclare une année d'arrivée 10 ans avant le recensement alors qu'il dit par ailleurs résider à l'étranger 5 ans avant le recensement.

Pour les EAR 2011 et 2012 il est incohérent de répondre habiter en France le 1<sup>er</sup> janvier de l'année précédente et être arrivé en France l'année précédente, à moins d'être arrivé le 1<sup>er</sup> janvier, ce qui doit sans doute concerner très peu de cas. Par contre, il est possible d'avoir résidé en France 5 ans avant le recensement et d'être arrivé en France ultérieurement, puisque la personne a alors eu le temps de partir s'installer à l'étranger au moins un an, puis de revenir en France, en moins de cinq années. Il s'agit alors d'un aller-retour. Pourtant lorsqu'on regarde plus précisément les dates d'arrivée de ceux qui sont codés comme habitant en France 5 ans avant le recensement, on constate que 36 % d'entre eux ont déclaré être entrés en France cette année-là, ce qui est incompatible avec l'hypothèse d'un aller-retour. Le tableau 1 indique la répartition des migrants selon leurs réponses aux deux questions. La non réponse à l'année d'entrée est supérieure à 20 % pour les EAR de 2004 à 2010 et atteint presque 25 % pour les EAR 2011 et 2012, ce qui montre qu'il y a une incertitude sur l'année d'entrée pour beaucoup de migrants. Néanmoins, comme le montre le tableau 1, ce sont surtout les migrants qui sont classés résidant en France 5 ans (ou 1 an) auparavant qui ne répondent pas à la question sur leur année d'entrée, ce qui n'a donc a priori pas d'impact sur l'estimation des flux

#### Encadré 1

##### LE CAS DES VISAS DE LONG SÉJOUR VALANT TITRES DE SÉJOURS (VLS-TS)

Depuis 2009 il est possible de séjourner en France muni de visas de long séjour valant titre de séjour (VLS-TS). Ils ont été créés par la circulaire interministérielle du 29 mars 2009 afin de faciliter les démarches administratives. Ils étaient au départ disponibles seulement pour certaines catégories : étudiants, visiteurs, salariés et conjoints de Français. Ces visas sont délivrés par les ambassades de France dans les pays tiers et ont une durée de validité comprise entre 6 mois et 1 an. Depuis 2009 il y a environ 100 000 demandes par an (Secrétariat général à l'immigration et à l'intégration, 2013), dont 60 000 étudiants. Une fois le visa accordé, le migrant doit écrire une lettre à l'Office français de l'immigration et de l'intégration (OFII) afin de passer une visite médicale pour que son visa de long séjour soit validé. Dans les faits, environ 90 000 personnes sur les 100 000 demandes écrivent à l'OFII et 85 000 passent effectivement la visite médicale (ils disposent de trois mois à cet effet, mais en réalité cela peut prendre plus d'un an). Une fois cette visite passée, on attribue un numéro d'étranger à l'immigrant qui est alors inscrit dans le fichier national des

étrangers (FNE). Lorsqu'un individu détenteur d'un VLS-TS passe dans une préfecture afin de renouveler son titre ou pour faire enregistrer un changement de résidence principale, alors il est enregistré dans la base AGDREF et est automatiquement supprimé du FNE. Par conséquent, la base AGDREF ne contient que les VLS-TS passés par une préfecture, qui sont les seuls susceptibles d'être restés au moins un an en France (car par définition ils sont entrés sur le territoire après le début de validité de leur visa), et donc d'être considérés comme des migrants de long terme. Depuis 2009, l'extraction des données pour constituer la base AGDREF se fait 15 mois après la fin de l'année dont on souhaite connaître les flux. Par exemple, la base constituée des demandes faites en 2011 est extraite en mars 2013. Ce délai permet donc de ne pas manquer des migrants arrivés en France avec un VLS-TS qu'ils renouvellent ensuite en préfecture. Environ 7 000 personnes en 2009 et 15 000 en 2010 ayant un VLS-TS ne sont pas passées par une préfecture. Parmi ces 22 000 titres, 12 000 ont une durée de validité d'un an (dont 9 100 étudiants).

constitués des migrants arrivés l'année précédant le recensement. Cela peut s'expliquer par le fait que certains d'entre eux, étant depuis longtemps en France, ne se souviennent plus exactement quand ils y sont entrés. Cependant, la variable du lieu de résidence antérieure est redressée dans le traitement des *EAR*, et il se peut qu'elle soit à l'origine des incohérences observées avec la variable d'année d'entrée qui, elle, n'est pas redressée<sup>9</sup>.

La règle retenue ici pour compléter les données manquantes et corriger les incohérences est la suivante. Si la résidence antérieure indique que la personne recensée en janvier d'une année *N* résidait à l'étranger, 5 ans (ou 1 an) auparavant, on garde les années d'entrée cohérentes avec cette information, à savoir les années d'entrée supérieures ou égales à *N*-5 (ou *N*-1). Pour les années d'entrée incohérentes avec la résidence antérieure à l'étranger ou non renseignées, on les redresse par hot-deck sur ceux qui étaient à l'étranger 5 ans (ou 1 an) auparavant et qui ont répondu de façon cohérente à l'année d'arrivée, en utilisant les variables de sexe, âge et nationalité. Si la résidence antérieure indique que la personne recensée était en France 5 ans (ou 1 an) auparavant, la méthode est légèrement différente. Si l'année d'arrivée est renseignée, on garde cette information, même en cas d'incohérence avec la résidence antérieure, car on considère que l'incohérence vient soit d'un mauvais redressement de la variable sur la résidence antérieure soit du fait que le migrant a effectué un aller-retour entre la France et l'étranger. On complète ensuite les données manquantes sur l'année d'arrivée, par hot-deck

sur ceux qui résidaient en France 5 ans (ou 1 an) auparavant et qui ont répondu à l'année d'entrée, en prenant toujours en compte le sexe, l'âge et la nationalité. Cette règle permet donc de calculer pour chaque individu recensé une année d'entrée cohérente avec sa résidence antérieure, tout en tenant compte du fait que certaines réponses sur la résidence antérieure ont pu être mal redressées et que certaines personnes ont pu faire des allers-retours entre la France et l'étranger.

Contrairement à la base des titres de séjour, les *EAR* sont restreintes au champ des migrants de long terme, mais elles contiennent en plus les migrants mineurs ainsi que les migrants de l'Espace économique européen et de la Suisse. De plus, le recensement peut inclure des personnes en situation illégale (« sans-papiers ») car n'ayant pas de titre de séjour. Les champs des *EAR* et de la base AGDREF bien que se recoupant, sont donc différents.

Il a déjà été souligné que les deux sources décrites plus haut fournissent des chiffres différents sur les flux migratoires (Cadenel et Menard, 2010) en raison de leurs spécificités propres. Afin de les comparer il est nécessaire de se positionner sur un champ identique. Le champ retenu est constitué des migrants de nationalité de pays tiers, nés à l'étranger, ayant

9. Le lieu de résidence antérieure est presque toujours redressé à « France » lorsqu'il n'est pas renseigné et ce redressement ne prend pas en compte l'année d'entrée si elle est renseignée (Brutel, 2014).

Tableau 1  
**Répartition des personnes étrangères nées à l'étranger selon leurs réponses brutes à l'année d'arrivée et les réponses imputées dans les *EAR* au lieu de résidence 5 ans (ou 1 an) avant**

				En %
Pays de résidence 5 ans auparavant	Arrivée il y a plus de 5 ans	Arrivée il y a moins de 5 ans	Non renseignée	Total
Étranger	0,5	15,5	2,3	18,3
France	55,3	6,8	19,6	81,7
Total	55,8	22,3	21,9	100
Pays de résidence un an auparavant	Arrivée il y a plus d'un an	Arrivée il y a moins d'un an	Non renseignée	Total
Étranger	0,2	3,2	0,6	4
France	70,2	1,8	24	96
Total	70,4	5	24,6	100

Lecture : les chiffres sont donnés en pourcentage et tiennent compte de la pondération des *EAR*. La table du haut porte sur les *EAR* de 2005 à 2010 et celle du bas sur les *EAR* 2011 et 2012 pour lesquelles le questionnaire a été modifié.

Champ : personnes nées à l'étranger de nationalité de pays tiers.

Source : *EAR* 2005-2012, Insee.

19 ans ou plus l'année de leur entrée<sup>10</sup> et résidant en France métropolitaine.

Sur ce champ, le nombre de premiers titres de séjour délivrés chaque année (157 000 en moyenne sur les années 2004-2012) est supérieur au nombre d'individus recensés comme arrivés l'année précédente (88 000 en moyenne sur les années 2004-2012).

Du fait qu'une partie des migrants obtiennent leur premier titre de séjour une ou plusieurs années après leur arrivée sur le territoire, le nombre de premiers titres délivrés une année n'est pas égal au nombre de premiers titres délivrés à ceux qui sont entrés cette même année. Par exemple parmi les titres délivrés en 2008<sup>11</sup>, 37 % concernent des migrants qui sont arrivés en 2007 ou avant. Mais ceci ne suffit pas à expliquer l'écart constaté avec le recensement car certains migrants arrivés en 2008 vont obtenir leur titre en 2009 ou après, et ne sont donc pas comptés dans les premiers titres délivrés en 2008. Ils seront comptés dans les premiers titres délivrés en 2009 et après. Ainsi, 26 % des migrants entrés en 2008 et ayant obtenu un titre entre 2008 et 2012 ont en fait obtenu leur premier titre entre 2009 et 2012.

Une autre source de divergence tient au fait qu'en théorie, seuls les migrants de long terme comme définis en introduction doivent être recensés alors que la base AGDREF contient des migrants de long terme et de court terme. Il faut donc distinguer ces deux types de migrants dans cette base.

## Déterminer les migrants de long terme

On souhaite dans cette section déterminer les migrants dans la base AGDREF dont la durée réelle de séjour en France est d'au moins 12 mois. La date d'arrivée sur le territoire est en général différente de la date de début de validité du titre qui peut se situer plusieurs mois ou années après. C'est pour cette raison qu'il faut bien distinguer entre l'année d'arrivée et l'année d'obtention du titre, qui ne sont pas toujours égales. Ainsi, certains migrants peuvent avoir séjourné plus d'un an en France (et donc être considérés comme des migrants de long terme) avant d'avoir obtenu leur premier titre.

Par ailleurs, comme le souligne Thierry (2001), tous ceux qui obtiennent un premier titre d'un

an ou plus ne vont pas obligatoirement rester effectivement plus d'un an sur le territoire. En effet, on peut par exemple se voir accorder un premier titre d'un an pour séjourner en France, mais quitter le pays avant la fin de validité du titre, au bout de 6 mois par exemple. De même on peut aussi avoir un premier titre de 6 mois, et le renouveler avec un deuxième titre d'une validité de 6 mois encore, et rester un an au total sur le territoire. On se distingue donc ici des travaux de l'Ined qui retiennent les migrants dont au moins un des titres obtenu à une durée de validité de plus d'un an (considérant alors l'année d'arrivée de ce migrant comme l'année d'obtention de ce titre d'une validité supérieure ou égale à un an) (Mazuy *et al.*, 2008).

Quels sont donc les migrants de long terme, c'est à dire les migrants qui résident en France 12 mois ou plus ? Pour pouvoir le déterminer, on répartit dans un premier temps les migrants en trois catégories différentes : les *courts*, les *inters*, et les *longs*. Dans un deuxième temps, on ventile les *inters* dont le temps de présence est incertain en migrants de long terme et migrants de court terme. On utilise pour cela la date d'entrée sur le territoire, la dernière date de présence sur le territoire<sup>12</sup>, et la date de fin de validité la plus récente<sup>13</sup> de chaque migrant. Ces dates sont connues au jour près, même si on peut penser qu'il y a des incertitudes sur la date d'entrée qui peut parfois être déclarative, sans document attestant d'une date précise<sup>14</sup>. La dernière date de présence est calculée en prenant en compte tous les types de documents disponibles pour chaque individu (titres de séjour et renouvellements de titres, mais aussi convocations, autorisations provisoires de séjour et récépissés). Chacun de ces documents indique une date de dépôt qui correspond en théorie à un passage du migrant en préfecture. Ces dates de dépôts témoignent donc de la présence du migrant sur le territoire, et la dernière date de présence enregistrée correspond donc pour chaque migrant à la date de dépôt la plus récente répertoriée dans AGDREF.

10. On élimine du champ les personnes âgées de 18 ans pour éviter les attributions de titres à leur majorité des personnes entrées mineures. Sur la période 2004-2012, il y a en moyenne 14 500 migrants chaque année qui obtiennent un premier titre l'année de leur 18 ans, alors qu'il n'y en a que 9 500 qui l'obtiennent l'année de leur 19 ans.

11. C'est-à-dire dont le début de validité se situe en 2008.

12. Cette date est définie ici comme la date du dernier passage en préfecture enregistré.

13. En cas de renouvellements, on considère donc la fin de validité du dernier renouvellement effectué.

14. En général cette date peut être connue à partir du visa qui a servi à entrer en France pour la première fois et n'est donc pas erronée.



Les trois catégories de migrants sont définies de la manière suivante :

- Migrants *courts* : ce sont les migrants qui sont restés strictement moins d'un an en France de façon légale. La différence entre la fin de validité de leur séjour (en tenant compte des renouvellements successifs) et la date d'entrée est strictement inférieure à un an.
- Migrants *longs* : ce sont les migrants dont on est sûr qu'ils sont restés plus d'un an en France. La différence entre la dernière date de présence enregistrée en préfecture et la date d'entrée est supérieure ou égale à un an.
- Migrants *inters* : ce sont les migrants qui ne sont ni courts, ni longs, c'est-à-dire que bien qu'ils aient pu rester légalement plus d'un an sur le territoire, on ne sait pas s'ils sont partis avant d'avoir résidé au moins un an en France, puisqu'ils ont pu quitter le territoire avant la fin de validité de leur titre.

Les effectifs correspondant à chacune de ces catégories sont donnés par le tableau 2.

Sur la période 2004-2010, les *inters* représentant 21 % des demandes de premiers titres sont constitués à 46 % de migrants pour motif familial, à 30 % d'étudiants, à 10 % de migrants pour motifs économiques, à 8 % pour motifs divers et à 6 % de motifs humanitaires.

Afin de s'assurer de ne pas manquer de migrants de long terme, on considérera par la suite tous les *inters* comme des *longs* sauf pour les cas particuliers suivants. Les « saisonniers », dont le motif est « économique », ne restent pas plus d'un an d'affilé sur le territoire par définition (cf. encadré 2) ; par conséquent ils sont considérés comme *courts*. Les « visiteurs » (au sein du motif « divers ») ainsi que les « étudiants » n'ont *a priori*, de par leur motif, pas vocation à s'installer en France. Le titre de séjour au motif « visiteur » ne donne d'ailleurs pas le droit d'exercer une activité professionnelle. On considère alors que les « visiteurs » ainsi que les « étudiants » *inters*, dont le premier titre de séjour a une durée de validité inférieure à 1 an et qui n'ont pas renouvelé ce titre, sont des migrants de court terme, et on les reclasse dans

Tableau 2

**Répartition des nouveaux détenteurs d'un premier titre de séjour entre les trois catégories court, inter et long en fonction de l'année de début de validité des premiers titres**

	2004	2005	2006	2007	2008	2009	2010	2011	2012
Court	4 853	5 349	4 876	4 571	5 094	8 545	11 068	11 034	55 179
Inter	35 433	35 760	33 486	32 248	34 768	30 664	32 043	61 446	54 620
Long	123 160	117 917	117 359	108 419	118 401	124 236	115 847	82 284	44 880

Champ : premiers titres de migrants des pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine.  
Source : AGDREF, DSED.

## Encadré 2

### CAS DES SAISONNIERS ET TEMPORAIRES

L'EAR est une photographie de la population du 15 janvier à fin février et ne capte pas les effets saisonniers. Les travailleurs saisonniers sont principalement présents pendant l'été : leur carte de séjour porte la mention « saisonnier » s'ils s'engagent à maintenir leur résidence habituelle hors de France et à ne pas passer plus de 6 mois en France sur 12 mois consécutifs. Ils ne répondent normalement donc pas au *Recensement de la population*. Quant aux temporaires, leur titre est associé à une activité salariale d'une durée déterminée inférieure à douze mois, ils ne deviennent a priori pas résidents (sauf en cas de renouvellement de titre par exemple).

Les saisonniers ou temporaires sont identifiés par le motif « Saisonnier ou temporaire » dans AGDREF. La

base AGDREF permet de distinguer les saisonniers des temporaires à partir de la référence réglementaire du titre.

La population des saisonniers et temporaires est atypique : elle est masculine à 97 % ; 94% des personnes en relevant obtiennent un titre de 3 ans (ce qui représente 57 % de l'ensemble des titres de 3 ans délivrés), et les 6 % restants sont détentrices d'un titre de 3 mois à 2 ans.

Les saisonniers ne peuvent donc pas être considérés comme des migrants de long terme, tandis que les temporaires peuvent l'être seulement s'ils renouvellent leur titre de séjour de façon à rester effectivement plus d'un an en France.

la catégorie *court*. On suppose en effet que ces étudiants sont venus pour une année scolaire qui en général dure moins de 12 mois et sont donc repartis avant de passer un an en France. 99 % des « étudiants » de la catégorie *inter* ont reçu un premier titre dont la durée de validité est inférieure ou égale à 1 an, et parmi eux 84 % n'ont pas renouvelé ce premier titre. De même, tous les « visiteurs » *inters* ont obtenu un premier titre d'une durée de validité d'un an ou moins, et 78 % d'entre eux n'ont pas renouvelé ce titre.

Le tableau 3 montre que certains migrants classés dans la catégorie *court* ont en fait un premier titre d'une durée de validité supérieure ou égale à un an. Ceci n'est pas contradictoire puisqu'il s'agit en réalité de migrants ayant obtenu un visa valant titre de séjour (VLS-TS) et qui sont entrés sur le territoire après le début de validité de leur premier titre. On remarque également que 82 % des migrants *longs* ont obtenu un premier titre d'une durée de validité égale à un an. La majorité des migrants *inters* (52 %) ont quant à eux obtenu un premier titre d'une validité strictement supérieure à un an (dont 71 % un titre d'une validité de 10 ans). Il y a donc une véritable incertitude sur le fait que ces migrants *inters* dont le premier titre est valable strictement plus d'un an sont effectivement restés un an ou non sur le territoire. La méthode retenue ici surestime donc probablement le nombre de migrants de long terme présents dans la base AGDREF.

Pour la dernière année disponible (2012), l'information sur les renouvellements éventuels des migrants n'est pas encore disponible,

ce qui fait que la plupart des migrants dont le début de validité de leur premier titre commence en 2012 sont dans la catégorie *inter* ou *court* (cf. tableau 2). Certains sont déjà dans la catégorie *long* puisqu'ils sont entrés au moins un an avant de déposer leur demande de premier titre obtenu. Dans la suite, on ne prendra donc pas en compte cette année. On se restreint donc à la période 2004-2011 pour l'analyse des flux migratoires.

Le tableau 4 donne alors le nombre de nouveaux migrants dans la base AGDREF, estimé en considérant qu'ils étaient des migrants de long terme, en fonction de l'année de début de validité de leur premier titre d'une part et en fonction de l'année d'entrée en France d'autre part. Le nombre de migrants de long terme, déterminé à partir de la base AGDREF, qui entrent chaque année sur le territoire, s'avère alors inférieur au nombre de premiers titres délivrés et ce nombre diminue (131 231 en 2004 et 101 236 en 2011). Cela est dû au fait qu'une partie des titres délivrés entre 2004 et 2011 concerne des entrées sur le territoire antérieures à 2004. À l'inverse on n'observe pas encore ceux qui sont entrés entre 2004 et 2011, mais qui n'obtiendront leur titre qu'après 2012.

Le nombre d'individus identifiés comme nouveaux migrants par les *EAR* est supérieur au nombre de migrants de long terme obtenant leur premier titre l'année de leur arrivée (cf. figure I), mais il reste inférieur au nombre total de premiers titres délivrés à des migrants supposés de long terme (qui ne sont donc pas forcément arrivés l'année où ils obtiennent leur premier titre).

Tableau 3

**Répartition de la durée du premier titre obtenu selon chaque catégorie de migrant**

En %

	De 0 à 364 jours	365 ou 366 jours	367 jours et plus	Total
Court	82	13	5	100
Inter	6	42	52	100
Long	7	82	11	100

Champ : premiers titres des migrants de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine.  
Source : AGDREF, DSED.

Tableau 4

**Répartition des migrants de long terme selon l'année de début de validité du premier titre et l'année d'entrée en France**

Début de validité du premier titre	2004	2005	2006	2007	2008	2009	2010	2011
	147 577	142 691	139 636	128 949	139 963	147 979	141 027	137 671
Année d'entrée en France	2004	2005	2006	2007	2008	2009	2010	2011
	131 231	120 661	108 838	107 582	117 529	114 464	109 909	101 236

Lecture : ce tableau donne le nombre de nouveaux détenteurs d'un premier titre de séjour considérés comme des migrants de long terme, après répartition de la catégorie *inter* en courts et longs, en fonction de l'année de début de validité du premier titre et de l'année d'entrée en France.

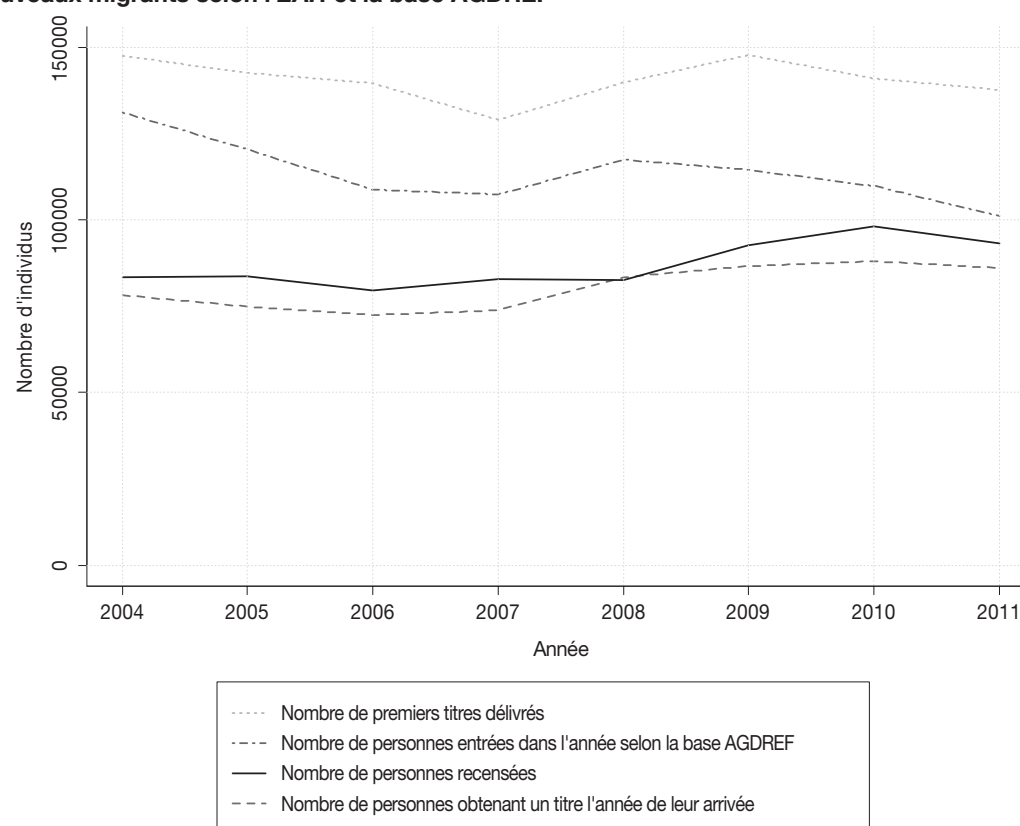
Champ : premiers titres de migrants des pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine, migrants de long terme.  
Source : AGDREF, DSED.

## Le modèle de réconciliation

On a restreint l'analyse à un champ commun aux *EAR* et à la base AGDREF.

La modélisation qui suit permet de prendre en compte le fait que certains migrants obtiennent leur premier titre de séjour plusieurs années après leur entrée sur le territoire,

Figure I  
Nouveaux migrants selon l'*EAR* et la base AGDREF



Lecture : en trait plein est indiqué le nombre de personnes recensées dans les *EAR* chaque année. Les traits discontinus au contraire concernent la base AGDREF : le nombre de premiers titres délivrés correspond au nombre de titre de séjour délivré une année à des migrants de long terme quelle que soit leur année d'arrivée, le nombre de personnes entrées dans l'année correspond au nombre de migrants de long terme entrée une année et ayant reçu un titre de séjour cette même année ou une année ultérieure, enfin le nombre de personnes obtenant un titre l'année de leur arrivée correspond au nombre de migrant de long terme entrant une année et obtenant leur premier titre de séjour cette même année.

Champ : migrants de long terme, de nationalité de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine. Source : AGDREF, DSED et *EAR*, Insee, 2005-2012.

Tableau 5  
Résultats des estimations avec  $\tau = 1$

	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_7$	$\alpha_8$	$\alpha_9$	$\alpha_{10}$	$\alpha_{11}$
Moyenne (%)	57,1	14,5	6,2	4,8	3,7	3,2	2,7	2,4	2	1,7	1,8
Écart-type (%)	0,04	0,03	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02

	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$	$\theta_6$	$\theta_7$	$\theta_8$	$\theta_9$	$\theta_{10}$	$\theta_{11}$
Moyenne (%)	58,7	53	52,4	49,4	48,8	47,5	44,2	40,6	37,1	35,4	32,5
Écart-type (%)	0,06	0,06	0,06	0,06	0,06	0,06	0,06	0,05	0,05	0,05	0,05

	$V_{2004}$	$V_{2005}$	$V_{2006}$	$V_{2007}$	$V_{2008}$	$V_{2009}$	$V_{2010}$	$V_{2011}$	$V_{2004}$
Moyenne	148 627	142 649	132 002	134 002	148 242	156 737	162 896	160 371	148 627
Écart-type	155	153	149	154	172	189	207	232	155

Lecture : les écart-types indiqués correspondent aux écart-types de la loi *a posteriori* estimés sur 100 000 itérations.

Champ : migrants de long terme, de nationalité de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine. Source : AGDREF, DSED et *EAR*, Insee, 2005-2012.

et que d'autres sont mal identifiés dans les *EAR*.

### Modélisation

L'idée sur laquelle repose cette modélisation est que les deux sources exploitées ici donnent une vision partielle mais complémentaire des vrais flux migratoires sous-jacents. Un modèle de *capture-recapture*, permettant d'estimer la taille d'une population en tirant deux échantillons indépendants au sein de cette population et en connaissant le nombre d'individus qui sont présents dans les deux échantillons à la fois, ne peut pas être mis en place, car cela supposerait de connaître parmi les personnes recensées celles qui ont un titre de séjour ou alors de connaître parmi les personnes qui ont un titre de séjour celles qui ont été recensées, ce qui n'est pas le cas ici.

Pour chaque année  $N$ , on considère la génération de migrants de long terme entrés en France cette année  $N$  et on note  $V_N$  ce nombre de migrants réellement entrés. De plus, on note  $E_{i,N}$  le nombre de migrants de long terme entrés l'année  $N$  et recensés au début de l'année  $N+i$  et  $A_{i,N}$  le nombre de migrants de long terme qui sont entrés l'année  $N$  et ont obtenu leur premier titre de séjour l'année  $N+i-1$ . Les  $E_{i,N}$  et les  $A_{i,N}$  sont connus (cf. annexe 2) alors que les  $V_N$  sont inconnus.

On suppose que chaque migrant de long terme entré l'année  $N$  a une probabilité  $\theta_{i,N}$  d'être recensé l'année  $N+i$  et une probabilité  $\alpha_{i,N}$  d'obtenir son premier titre de séjour l'année  $N+i-1$ . On a donc :

$$\begin{cases} E_{i,N} & \sim \text{Binom}(\theta_{i,N}, V_N) \\ A_{i,N} & \sim \text{Binom}(\alpha_{i,N}, V_N) \end{cases}$$

Les taux  $\theta_{i,N}$  sont associés à la *non-présence* aux *Enquêtes annuelles de recensement* par la différence  $1-\theta_{i,N}$ , ce qui est différent de la *non-réponse* au recensement de la population, qui désigne les formulaires adressés dans les logements et non remplis. Cette non-réponse là est de l'ordre de 3 % pour chaque *EAR*, et elle est corrigée en amont dans le traitement des *EAR*.

La part de *non-présents* au recensement recouvre quant à elle plusieurs raisons : sorties du territoire, décès, acquisition de la nationalité (car les migrants sont par définition de nationalité étrangère), mauvaise identification comme nouveau migrant au recensement ou bien non recensement.

Certains migrants de long terme ont pu penser qu'ils ne s'installeraient pas en France plus d'un an, et n'ont donc pas répondu au recensement lors de leur première année passée en France. Certains migrants de long terme ont également pu ne pas être identifiés comme nouveaux migrants l'année de leur arrivée en raison d'une mauvaise imputation de leur année d'entrée.

Par ailleurs, les taux de présence au recensement ainsi déterminés sont relatifs aux estimations des  $V_N$  qui dépendent en particulier des  $A_{i,N}$ . Or les  $A_{i,N}$  qui représentent le nombre de migrants de long terme entrés une année  $N$  et obtenant leur titre de séjour la même année, sont entachés d'incertitude puisque 16 % de ceux qu'on a considéré comme *longs* sont en fait des *inters* et qu'une partie d'entre eux a pu repartir avant de rester un an en France. Donc on surestime probablement le nombre  $V_N$  de migrants de long terme à partir de la base AGDREF. Dans ce contexte, les  $\theta_{i,N}$  peuvent être sous-estimés.

On suppose par la suite que les  $\theta_{i,N}$  ainsi que les  $\alpha_{i,N}$  ne dépendent pas de l'année  $N$ . Cette hypothèse sous-entend que le comportement des migrants est le même quelle que soit leur année d'arrivée. Cette hypothèse est justifiée par le fait que les rapports  $A_{i+1,N}/A_{i,N}$  et  $E_{i+1,N}/E_{i,N}$  ne semblent pas en première approximation dépendre de l'année  $N$  pour les années considérées (cf. figure II). Néanmoins, à partir de 2009, les rapports  $A_{i+1,N}/A_{i,N}$  semblent se modifier : il y a moins de migrants qui demandent leur titre l'année après leur année d'arrivée. Ce constat est à relier au fait qu'à partir de 2009, certains migrants ont la possibilité de demander des VLS-TS, ce qui raccourcit les délais d'obtention de titre.

On considère alors le modèle suivant :

$$\text{Modèle A : } \begin{cases} E_{i,N} & \sim \text{Binom}(\theta_i, V_N) \\ A_{i,N} & \sim \text{Binom}(\alpha_i, V_N) \end{cases}$$

Les données peuvent être ventilées par sexe, âge à l'entrée en France ou nationalité. Si l'on considère la population totale comme étant composée de  $n$  sous-populations distinctes, on peut également considérer le modèle suivant :

$$\text{Modèle B : } \begin{cases} E_{i,N}^j & \sim \text{Binom}(\theta_i^j, V_N^j) \\ A_{i,N}^j & \sim \text{Binom}(\alpha_i^j, V_N^j) \end{cases}$$

où l'indice  $j$  ( $1 \leq j \leq n$ ) fait référence à la  $j^{\text{e}}$  sous-population. On obtient ainsi des paramètres sur les  $n$  sous-populations.

On obtient alors le paramètre global de plusieurs sous-populations en calculant la moyenne des paramètres de chacune des sous-populations  $j$ , pondérée par le total de la sous-population en question. Ce total est précisément estimé par le modèle pour plusieurs années  $N$ , à savoir les  $V_N^j$ . On utilise comme poids dans la moyenne pondérée la somme de ces  $V_N^j$  sur l'ensemble des années disponibles (2004-2011).

### Méthode d'estimation

Les paramètres de ces modèles seront estimés par inférence bayésienne (cf. encadré 3)<sup>15</sup>. Comme il n'y a pas d'autres sources de données, il y a très peu d'information *a priori* sur

les paramètres. C'est pourquoi les lois *a priori* choisies sont des lois uniformes :

$$\alpha_i \sim \text{Unif}(0,1)$$

$$\theta_i \sim \text{Unif}(0,1)$$

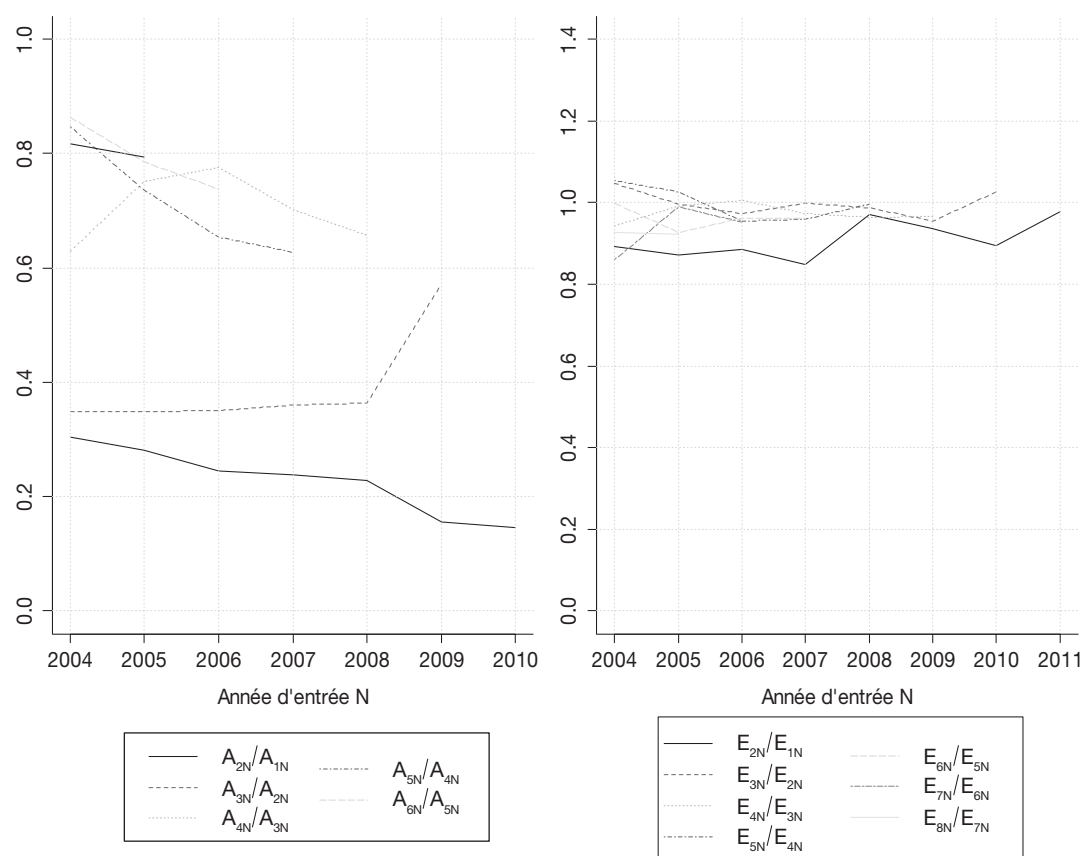
$$V_N \sim \text{Unif}(V_{\text{inf}}, V_{\text{sup}})$$

On sait déjà que les  $V_N$  sont supérieurs aux  $E_{i,N}$  et aux  $A_{i,N}$ , c'est pourquoi on choisit comme borne inférieure de l'*a priori* sur les  $V_N$ ,  $V_{\text{inf}}=80\,000$ . Par contre on ne sait rien de la borne supérieure.

On peut choisir une borne supérieure très grande de façon à ce que les vraies entrées

15. Les lois *a posteriori* ont été simulées en utilisant un algorithme de Monte Carlo par chaînes de Markov à l'aide du logiciel JAGS (Just Another Gibbs Sampler) et du logiciel R.

Figure II  
Rapports  $A_{i+1,N}/A_{i,N}$  et  $E_{i+1,N}/E_{i,N}$  pour différentes valeurs de  $i$  en fonction de l'année  $N$



Lecture : la stabilité apparente de ces rapports suggère que les  $\alpha_{i+1,N}/\alpha_{i,N}$  et  $\theta_{i+1,N}/\theta_{i,N}$  ne dépendent pas de l'année  $N$ , ce qui laisse penser que les taux  $\alpha_{i,N}$  et  $\theta_{i,N}$  ne dépendent pas de  $N$ .  
Champ : migrants de long terme, de nationalité de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine.  
Source : AGDREF, DSED et EAR, Insee, 2005-2012.



soient inférieures à cette borne. Pour la suite des estimations, on retiendra donc (sauf mention du contraire)  $V_{\text{sup}} = 500\,000$ .

En l'absence d'hypothèses supplémentaires, le modèle n'est pas identifiable. Il faut donc ajouter une contrainte sur les paramètres. On sait déjà que  $\sum_{i \geq 1} \alpha_i \leq 1$  car c'est la proportion de la population totale entrante qui finira par avoir un titre un jour ou l'autre. Au contraire,  $1 - \sum_{i \geq 1} \alpha_i$  est la part des migrants de long terme entrés une année  $N$  qui n'auront jamais de titre. Cela ne veut pas dire que ces migrants qui n'obtiennent pas de titre s'installent de manière permanente et illégale en France puisqu'ils ont pu décéder, sortir du pays ou même être naturalisés avant d'avoir eu la possibilité d'obtenir leur premier titre. Par la suite, on notera  $\tau = \sum_{i=1}^{11} \alpha_i$  la proportion moyenne des migrants arrivant chaque année

qui obtiendront un titre dans les 10 années suivant leur arrivée sur le territoire. Certains migrants peuvent obtenir leur premier titre plus de 10 années après leur arrivée, mais cela concerne très peu de cas puisqu'il est possible pour les migrants d'être régularisé après dix ans de séjour sur le territoire (Bègue, 2007). Parmi ceux qui ont obtenu leur premier titre sur les années 2004-2011, seulement 2,5 % sont entrés plus de 10 ans avant l'année de leur obtention de titre. Imposer  $\tau \leq 1$  ne suffit pas à avoir de bonnes estimations. En fait il est nécessaire d'imposer la valeur de  $\tau$  pour que le modèle soit identifiable. En effet, si on ne contraint pas la valeur de  $\tau$ , les distributions *a posteriori* des  $V_N$  ressembleront beaucoup aux distributions *a priori* (cf. encadré 3 et annexe 1).

On peut donc établir différents scénarios en fonction des valeurs de  $\tau$  retenues. Le scénario le plus simple est de considérer que  $\tau = 1$ , c'est-à-dire que chaque migrant entré en France va finir par obtenir un premier titre de séjour avant de sortir du champ et avant d'avoir passé 11 ans sur le

### Encadré 3

#### PRINCIPE DE L'INFÉRENCE BAYÉSIENNE

Comme dans la plupart des modèles statistiques, on dispose de  $n$  observations  $x = (x_1, \dots, x_n)$  qu'on suppose générées selon un modèle paramétrique de loi  $P(\cdot, \theta)$ , où  $\theta$  représente les paramètres. Dans cet article, les observations sont les  $E_{i,N}$  (personnes présentes au recensement) et les  $A_{i,N}$  (personnes obtenant un titre de séjour) et les paramètres sont les  $\theta_i$  (taux de présence au recensement), les  $\alpha_i$  (taux d'obtention de titre de séjour) et les  $V_N$  (flux de migrants de long terme).

On souhaite estimer les paramètres à l'aide des observations. Traditionnellement, on construit pour cela un estimateur qui est une fonction qui associe à chaque jeu d'observations donné une valeur des paramètres. Par exemple, un estimateur très utilisé est l'estimateur du maximum de vraisemblance.

L'approche bayésienne diffère en ce qu'elle considère les paramètres  $\theta$  comme des variables aléatoires, et cherche à **estimer la distribution de  $\theta$  en fonction des observations  $x$** , et non pas seulement une valeur ponctuelle. Cette distribution est donnée par la formule de Bayes qui inverse les probabilités conditionnelles :

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

La **densité *a posteriori*** de  $\theta$ ,  $P(\theta|x)$ , est donc proportionnelle à la vraisemblance du modèle,  $P(x|\theta)$ ,

multipliée par la **densité *a priori***  $P(\theta)$ . L'inférence bayésienne consiste ainsi à actualiser l'information sur les paramètres  $\theta$  au travers du prisme des observations. On peut obtenir un estimateur ponctuel des paramètres via l'espérance de la distribution *a posteriori* :

$$\hat{\theta} = E[\theta|x]$$

Dans l'approche bayésienne, outre le modèle il faut donc également spécifier la densité *a priori* des paramètres. Cela représente l'information que l'on a sur les paramètres avant toute observation. Un choix possible de lois qui modélisent des *a priori* non informatifs sont les lois uniformes qui donnent un poids égal à chaque valeur possible du paramètre.

Il n'est pas toujours possible de calculer explicitement la distribution *a posteriori*, ce qui a longtemps limité la mise en œuvre pratique de la méthodologie bayésienne. Mais, grâce à des progrès à la fois méthodologiques et computationnels, il est maintenant possible d'évaluer ou d'échantillonner la plupart de ces distributions. Les algorithmes utilisent la méthode de Monte-Carlo par chaîne de Markov (MCMC pour *Markov chain Monte Carlo*). Ils génèrent des chaînes de Markov dont la loi stationnaire correspond à la distribution *a posteriori* des paramètres  $\theta$ . On a ici utilisé le programme JAGS (Just Another Gibbs Sampler) qui permet, à l'aide du logiciel R, de générer de telles chaînes de Markov.

territoire. On considère aussi le cas où  $\tau = 0,8$ , ce qui correspond à la situation où 20 % des migrants de long terme entrant une année n'obtiendront pas de titre de séjour dans les dix années qui suivent l'année de leur entrée.

## Résultats et interprétations

### Résultats du modèle agrégé A

Les tableaux 5 et 6 donnent les estimations des paramètres pour le modèle agrégé avec  $\tau = 1$  et  $\tau = 0,8$  respectivement. Le cas  $\tau = 1$  donne une estimation plancher des flux entrants de migrants  $V_N$  car on considère alors que tous les migrants de long terme obtiennent leur premier titre de séjour dans les dix premières années après leur entrée en France. Or il est possible que certains migrants repartent sans avoir obtenu de titre de séjour ou qu'ils l'obtiennent après dix ans de séjour. Ainsi on peut déterminer qu'il y avait au moins 148 000<sup>16</sup> entrées en 2004 selon le modèle A. Toujours selon ce modèle, la première année de leur entrée, seulement 59 % des migrants de long terme sont présents au recensement, puis ce chiffre diminue avec le nombre d'années passées en France, ce qui s'explique par les sorties du champ (sorties du territoire, acquisitions de la nationalité française ou décès). Dès la troisième année, cette diminution semble se stabiliser avec un taux de recensement qui diminue en moyenne de 2,5 % par an.

En moyenne, plus d'un migrant sur deux (57 %) reçoit son premier titre de séjour l'année de son entrée sur le territoire et plus de 70 %

avant la fin de l'année suivante. Néanmoins, on remarque que les obtentions de titres peuvent se faire plusieurs années après l'entrée pour une part non négligeable de la population entrante. La 10<sup>e</sup> année après l'entrée en France, encore 1,8 % de la population initiale obtient un titre.

Les estimations changent lorsqu'on impose  $\tau = 0,8$  au lieu de  $\tau = 1$ . Comme on peut s'y attendre, les estimations des  $\alpha_i$  et des  $\theta_i$  sont alors plus faibles et les estimations des  $V_N$  plus importantes. Ce qui ne change pas, ce sont les rapports  $\alpha_{i+1}/\alpha_i$  et  $\theta_{i+1}/\theta_i$ . L'incertitude sur  $\tau$  se répercute donc dans l'incertitude sur les niveaux des paramètres, mais pas sur leurs écarts relatifs. On rappelle qu'une autre incertitude plane sur le nombre de migrants de long terme identifiés dans la base AGDREF, du fait de sorties possibles des *inters* avant qu'ils n'aient passé un an en France.

Si l'on assimile les taux  $\theta_i$  à des « taux de survie », on peut alors déterminer des risques instantanés correspondant à la probabilité instantanée de quitter le champ, définis par :

$$h_i = \frac{\theta_i - \theta_{i+1}}{\theta_i}$$

où  $i$  indique l'année. La figure III compare les risques instantanés calculés dans le cas  $\tau = 1$  et le cas  $\tau = 0,8$ . Ces taux sont très proches et suivent la même évolution au cours du temps puisqu'ils diminuent de la première à

16. Toutefois, il ne faut pas oublier que le nombre de migrants de long terme identifiés dans la base AGDREF est peut-être surestimé, ce qui fait que ce chiffre présenté comme borne inférieure ne l'est pas tout à fait : il est lui-même surestimé.

Tableau 6  
Résultats des estimations avec  $\tau = 0,8$

	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_7$	$\alpha_8$	$\alpha_9$	$\alpha_{10}$	$\alpha_{11}$
Moyenne (%)	46,2	11,4	4,8	3,8	2,9	2,5	2,2	1,9	1,6	1,3	1,4
Écart-type (%)	0,03	0,03	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$	$\theta_6$	$\theta_7$	$\theta_8$	$\theta_9$	$\theta_{10}$	$\theta_{11}$
Moyenne (%)	48,4	43,7	43,3	41	40,7	39,6	37	34,4	31,8	30,7	28,6
Écart-type (%)	0,06	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05
	$V_{2004}$	$V_{2005}$	$V_{2006}$	$V_{2007}$	$V_{2008}$	$V_{2009}$	$V_{2010}$	$V_{2011}$	$V_{2012}$	$V_{2013}$	$V_{2014}$
Moyenne	179 624	172 661	160 908	162 414	180 140	190 502	198 219	195 837	179 624		
Écart-type	200	200	195	203	229	251	278	313	200		

Lecture : les écart-types indiqués correspondent aux écart-types de la loi a posteriori estimés sur 100 000 itérations.

Champ : migrants de long terme, de nationalité de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine.  
Source : AGDREF, DSED et EAR, Insee, 2005-2012.

la troisième année, puis augmentent jusqu'à la septième année et semblent ensuite se stabiliser, suivant une courbe en « U ».

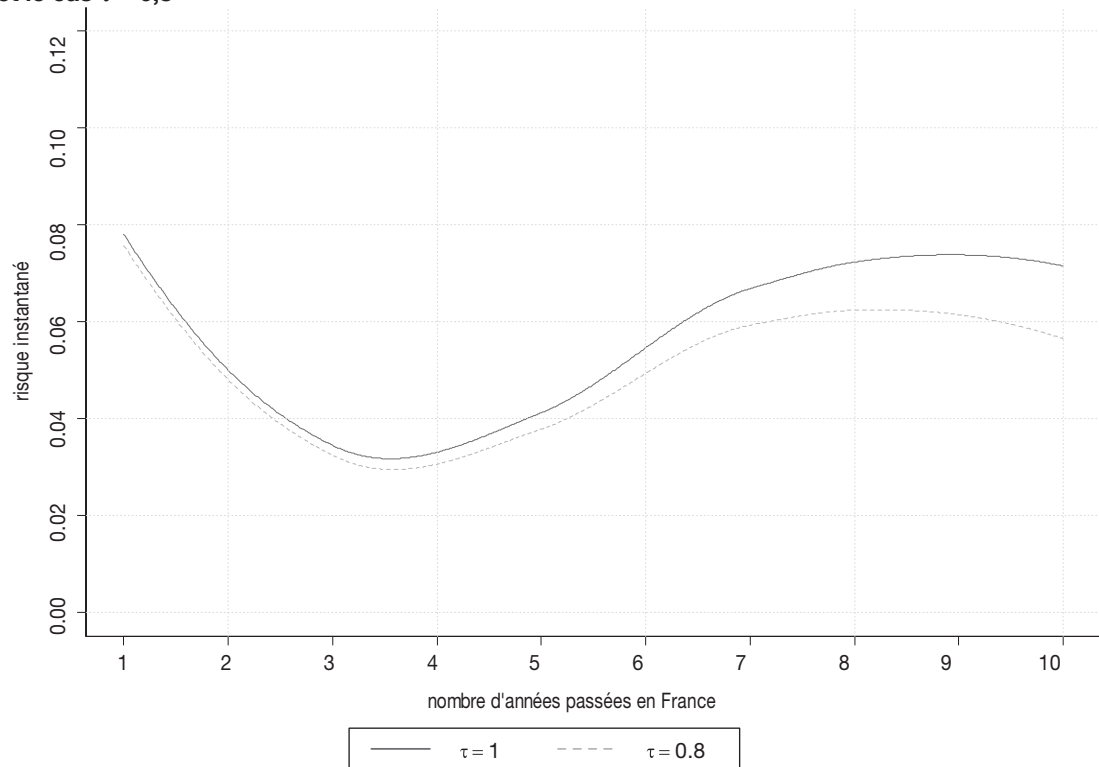
Comme le montre la figure IV, les estimations des flux entrant semblent suivre la même évolution que les flux détectés par le recensement, c'est-à-dire qu'on observe une légère diminution des flux entre les années 2004 et 2006 suivie par une augmentation entre 2006 et 2010 (18 600 entrées de plus selon les *EAR* et 30 100 selon l'estimation des flux avec  $\tau = 1$ ) et une très légère diminution entre 2010 et 2011. Pour les quatre premières années, les estimations dans le cas  $\tau = 1$  correspondent aux nombres de premiers titres délivrés, mais par la suite elles deviennent plus grandes, de sorte que le nombre de premiers titres délivrés ne semble pas toujours fournir une bonne estimation de l'évolution des flux au cours du temps. Ceci provient du fait que l'évolution du nombre de premiers titres délivrés suit l'évolution du nombre d'entrées réelles  $V_N$  avec un décalage dans le temps, car les premiers titres délivrés une année  $N$  dépendent aussi des entrées

des années  $N - k$  ( $k = 1, 2, \dots$ ). De plus, à partir de 2009 vient s'ajouter un autre effet dû à l'apparition de la possibilité de demander des VLS-TS. Ces visas facilitent les demandes de titres, et sont attribués avant l'entrée sur le territoire ; par conséquent, le rapport  $A_{2,N}/A_{1,N}$  du nombre de personnes obtenant leur premier titre l'année après leur arrivée sur le nombre de personnes obtenant leur premier titre l'année de leur arrivée diminue à partir de  $N = 2009$ . L'hypothèse de la constance des  $\alpha_i$  ne semble donc plus vérifiée.

### Résultats par sous-population

Le modèle B a ensuite été estimé sur plusieurs sous-populations constituées de migrants ayant des caractéristiques communes. On a ainsi distingué les migrants selon le continent d'origine (Afrique, Amérique, Asie, Europe ou Océanie), selon le sexe, la catégorie d'âge, et enfin selon la nationalité en s'intéressant aux 15 nationalités les plus représentées dans la base des titres de séjour (sur le champ de l'étude).

Figure III  
Risque instantané  $h_i$  en fonction du nombre d'années  $i$  passées en France, pour le cas  $\tau = 1$  et le cas  $\tau = 0,8$



Lecture : les courbes présentées sont obtenues par régression polynomiale des risques instantanés  $h_i$  de sortir du champ étudié dans le cas où  $\tau = 1$  et dans le cas où  $\tau = 0.8$ . On remarque que ce risque n'est pas constant au cours du temps passé en France mais a une forme de « U ».

Champ : migrants de long terme, de nationalité de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine. Source : AGDREF, DSED et EAR, Insee, 2005-2012.

Dans les estimations qui suivent, on a considéré que  $\tau = 1$ . Il faut alors comparer, entre les sous-populations, les paramètres estimés avec précaution, car rien n'indique que les  $\tau$  de ces différentes sous-populations sont effectivement égaux entre eux. De plus, la proportion d'*inters* considérés à tort comme des *longs* varie certainement d'une sous-population à l'autre, ce qui rend délicate une comparaison en niveau entre les paramètres des sous-populations. Néanmoins, on sera parfois amené à comparer certaines sous-population entre elles : on supposera donc que les  $\tau$  correspondant sont similaires.

### Par continent d'origine

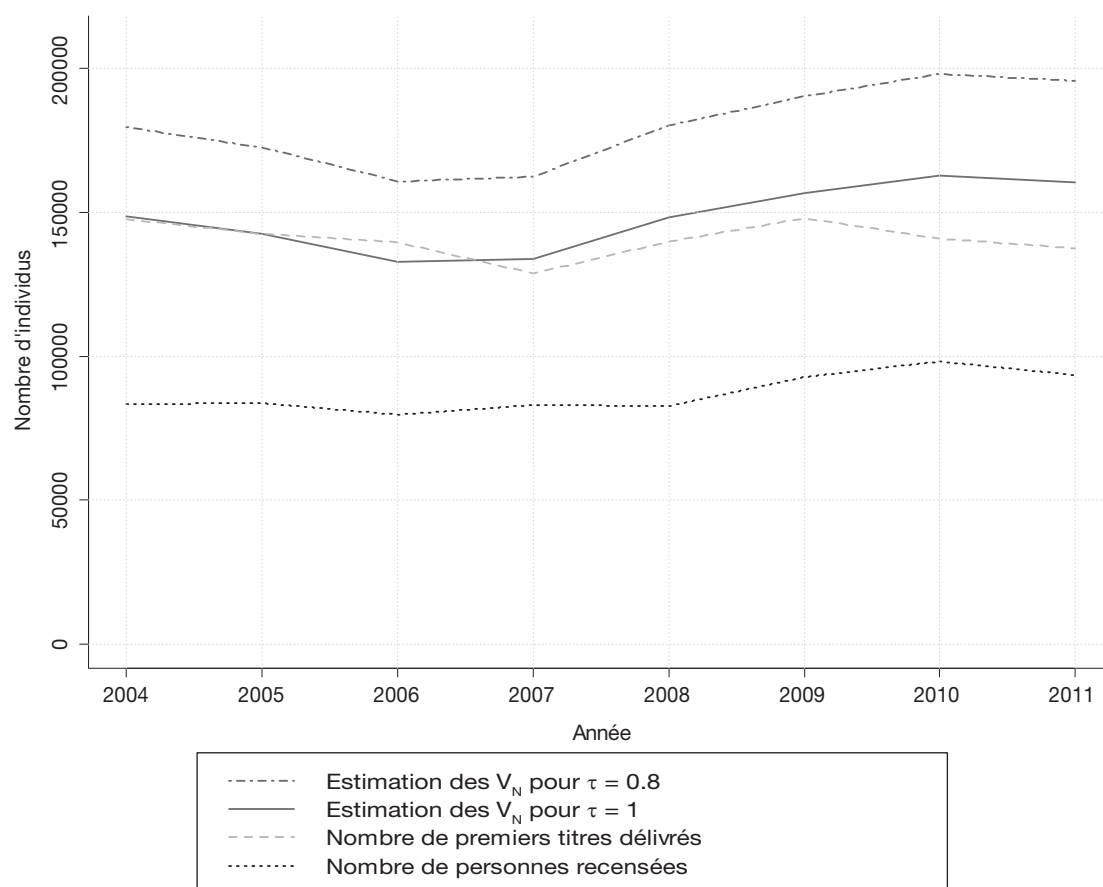
Les comportements vis-à-vis du recensement et des demandes de titre varient selon la provenance géographique du migrant<sup>17</sup> (cf. figure V). Exception faite des européens, plus d'un migrant sur deux obtient son premier titre de séjour l'année d'arrivée

sur le territoire. Au bout de dix années passées en France, les taux d'obtention de titre sont pratiquement les mêmes pour tous, et concernent une très faible part de la population initiale (environ 1 %).

En ce qui concerne le recensement, ce sont les migrants d'Amérique et d'Océanie qui sont les plus présents dans les *EAR*, avec un taux de présence de plus de 96 %. Au contraire, les Africains ne sont que 47 % à être présents au recensement la première année. Pour tous les continents sauf l'Afrique, on observe une chute du nombre de migrants recensés entre la première et la deuxième année de présence en France, ce qui traduit sans doute des sorties massives du territoire au bout d'un an. Les migrants d'Amérique et d'Océanie, qui ont des taux de sorties au bout d'un an ( $\theta_1 - \theta_2$ )

17. Le migrant ne résidait pas forcément sur le continent en question avant d'arriver en France, mais il a une nationalité correspondant à un pays de ce continent.

Figure IV  
Comparaison entre le nombre de personnes recensées, le nombre de premiers titres délivrés et les estimations des flux d'entrées pour  $\tau = 0,8$  et  $\tau = 1$ .



Champ : migrants de long terme, de nationalité de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine.

Source : AGDREF, DSED et EAR, Insee, 2005-2012.

les plus élevés, sont également ceux qui invoquent le plus souvent le motif « économique » (20 % et 37 % respectivement). Les différences  $\theta_i - \theta_{i+1}$  peuvent être assimilées à la différence entre les entrées et les sorties dans le recensement, mais ne correspondent pas forcément à des sorties du territoire. Pour cette raison, on peut avoir  $\theta_{i+1} > \theta_i$ , dans le cas où des migrants qui n'étaient pas présents au recensement  $i$  années après leur arrivée en France deviennent présents au recensement de l'année suivante ( $i + 1$ ). L'Amérique et l'Asie sont les continents qui ont la plus grande proportion d'étudiants parmi leurs migrants (27 % et 36 % respectivement). Le motif prédominant chez les migrants originaires d'Afrique est le motif « familial » (64 %), contrairement à ceux en provenance des autres continents (40 % ou moins).

À partir des paramètres estimés sur chaque sous-population, on peut calculer des paramètres globaux (caractéristiques de la population totale) comme indiqué plus haut (estimation du modèle). On retrouve alors presque exactement

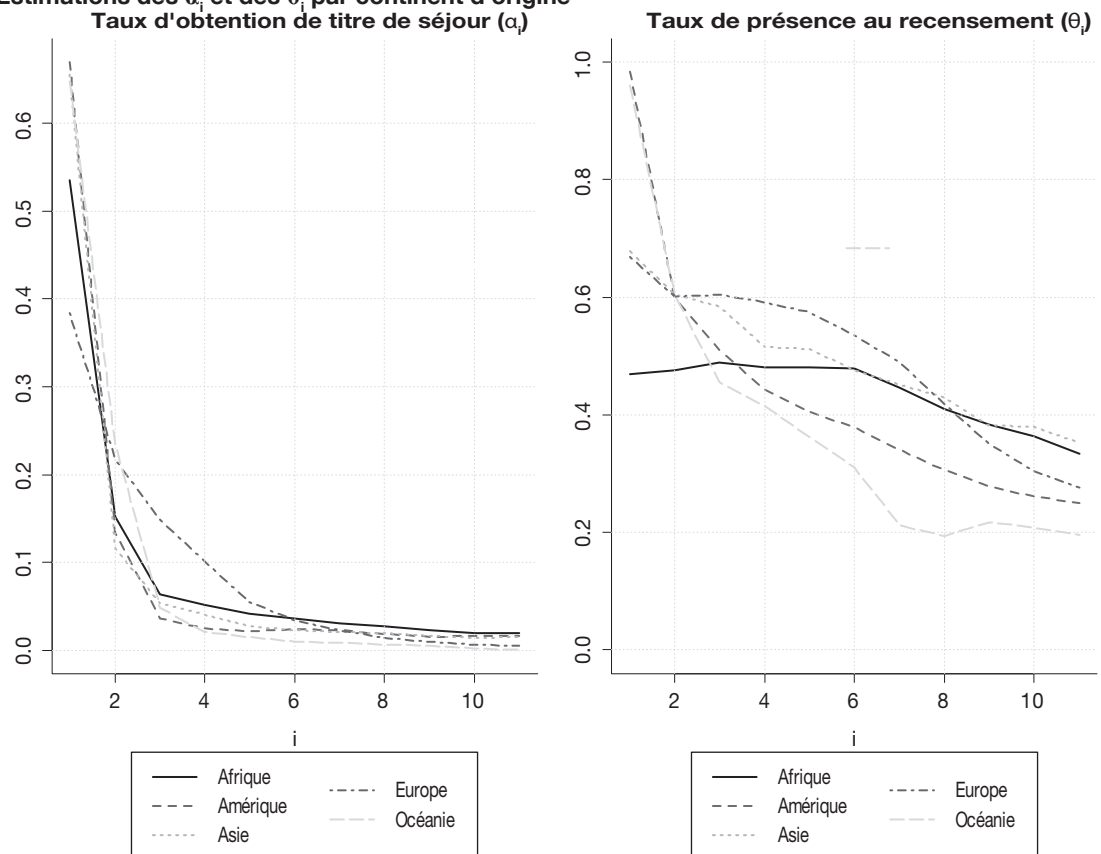
les mêmes paramètres que lors de l'estimation directe par le modèle A (cf. tableau 7).

### Par sexe et catégorie d'âge

Même si les hommes obtiennent leur premier titre presque en même temps que les femmes, ils sont moins présents au recensement que les femmes sur les 10 premières années passées en France (52 % la première année contre 61 % pour les femmes). Le fait que les taux de recensements convergent avec le temps indique soit qu'il y a plus de sorties de femmes que d'hommes, soit que les taux de réponse au recensement s'égalisent. Dans ce dernier cas, cela peut indiquer que les comportements des migrants vis-à-vis du recensement évoluent avec le temps pour devenir les mêmes quel que soit le sexe ou bien que ne restent en France que ceux qui ont la même attitude à l'égard du recensement.

La catégorie des plus jeunes (19-24 ans) a un meilleur taux de présence au recensement la

Figure V  
Estimations des  $\alpha_i$  et des  $\theta_i$  par continent d'origine



Lecture : ces courbes sont obtenues en estimant le modèle A séparément sur chaque sous-population avec  $\tau = 1$ .  
Champ : migrants de long terme, de nationalité de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine.  
Source : AGDREF, DSED et EAR, Insee, 2005-2012.



première année que les plus âgés (plus de 60 ans) (cf. figure VI). Cette catégorie, qui comporte la plus grande proportion d'étudiants (45 %),

présente la plus forte diminution du taux de présence sur les deux premières années de recensement. On peut supposer que cette diminution est

Tableau 7

**Estimations des paramètres globaux obtenues par moyenne pondérée des paramètres estimés par le modèle B pour chaque continent avec  $\tau = 1$**

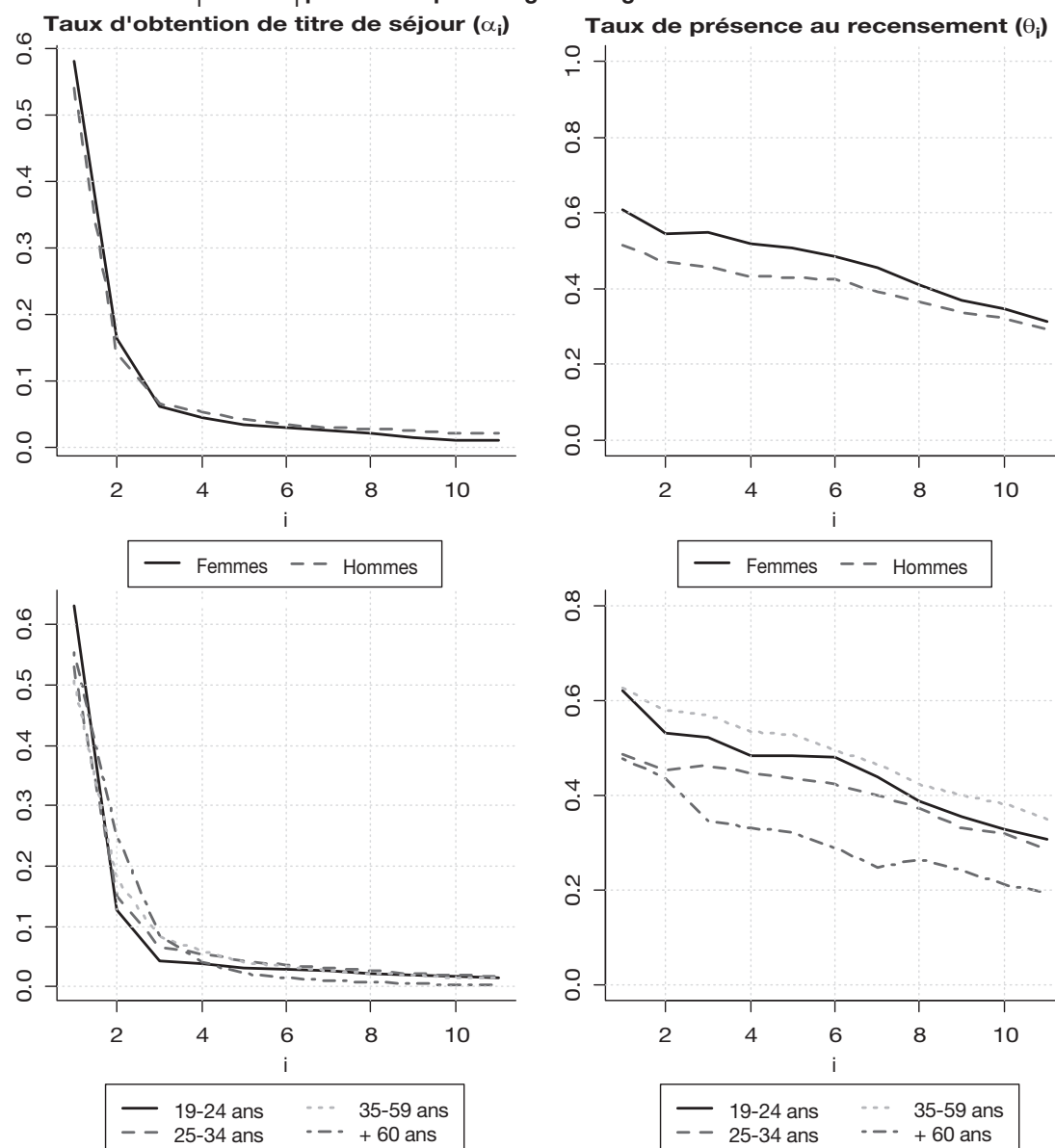
i	1	2	3	4	5	6	7	8	9	10	11
$\theta_i$	58,3	52,6	52,1	49,2	48,7	47,2	43,9	40,5	37	35,4	32,6
$\alpha_i$	56,7	14,6	6,4	5	3,7	3,2	2,7	2,3	2	1,7	1,8

Lecture : après avoir estimé le modèle séparément pour les cinq continents on obtient cinq séries de paramètres  $\theta_i$  et  $\alpha_i$ . Pour chaque année  $i$  on calcule alors la moyenne de ces cinq paramètres pondérée par le flux estimé moyen de migrants de chaque continent entrants entre 2004 et 2011 ( $V_{2004}$  à  $V_{2011}$ ).

Champ : migrants de long terme, de nationalité de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine. Source : AGDREF, DSED et EAR, Insee, 2005-2012.

Figure VI

**Estimations des  $\alpha_i$  et des  $\theta_i$  par sexe et par catégorie d'âge**



Lecture : ces courbes sont obtenues en estimant le modèle A séparément sur chaque sous-population avec  $\tau = 1$ .

Champ : migrants de long terme, de nationalité de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine. Source : AGDREF, DSED et EAR, Insee, 2005-2012.

en grande partie due à des départs d'étudiants qui ne seraient venu en France que pour un an ou deux. À l'exception des 25-34 ans, la diminution du taux de présence au recensement se fait par palier (ce qui est particulièrement visible pour les 19-24 ans avec une nette diminution entre la première et la deuxième année, entre la troisième et la quatrième année et entre la sixième et la septième année). À partir de la huitième année, la diminution du taux de présence au recensement semble s'égaliser pour chaque année et chaque catégorie d'âge (entre 2 % et 4 %).

Le taux d'obtention de titre la première année est le plus important pour les 19-24 et les 35-59 ans (plus de 62 %). Pour les autres catégories d'âge il se situe autour de 48 %.

### Par nationalité

Les taux de présence au recensement présentent de fortes disparités selon la nationalité, ce qui peut révéler des différences de comportement vis-à-vis du recensement. Parmi les 15 nationalités les plus représentées dans la base AGDREF (sur le champ considéré), une est européenne (russe), deux sont américaines (États-Unis et Brésil), trois sont asiatiques (turque, chinoise et japonaise), et les autres sont africaines (algérienne, marocaine, tunisienne, camerounaise, sénégalaise, ivoirienne, malienne, ex-zaïroise et congolaise).

D'après le modèle, la plupart des nationalités présentent une courbe concave pour l'évolution des taux de présence au recensement en fonction du temps, ce qui traduit un changement d'attitude des migrants envers le recensement (cf. figure VII). Le taux de présence au recensement des ex-Zaïrois par exemple passe de 26 % la première année à 40 % la cinquième année. De même, ce taux passe de 27 % à 43 % en l'espace de six ans pour les migrants de nationalité malienne. Parmi ces pays ayant une courbe concave, certains comme le Maroc, la Turquie ou le Sénégal marquent néanmoins une diminution du taux de présence au recensement entre la première et la deuxième année, ce qui est le signe de départs du territoire entre la première et la deuxième année.

Au contraire, d'autres pays comme la Chine, la Russie, les États-Unis, le Japon ou le Brésil ont une courbe toujours décroissante des taux de présence au recensement. Cette tendance est particulièrement marquée pour les États-Unis dont le taux de présence au recensement passe de 99 % à 20 % en sept ans.

À partir de la huitième année, les taux de présence au recensement diminuent en moyenne de 2 %

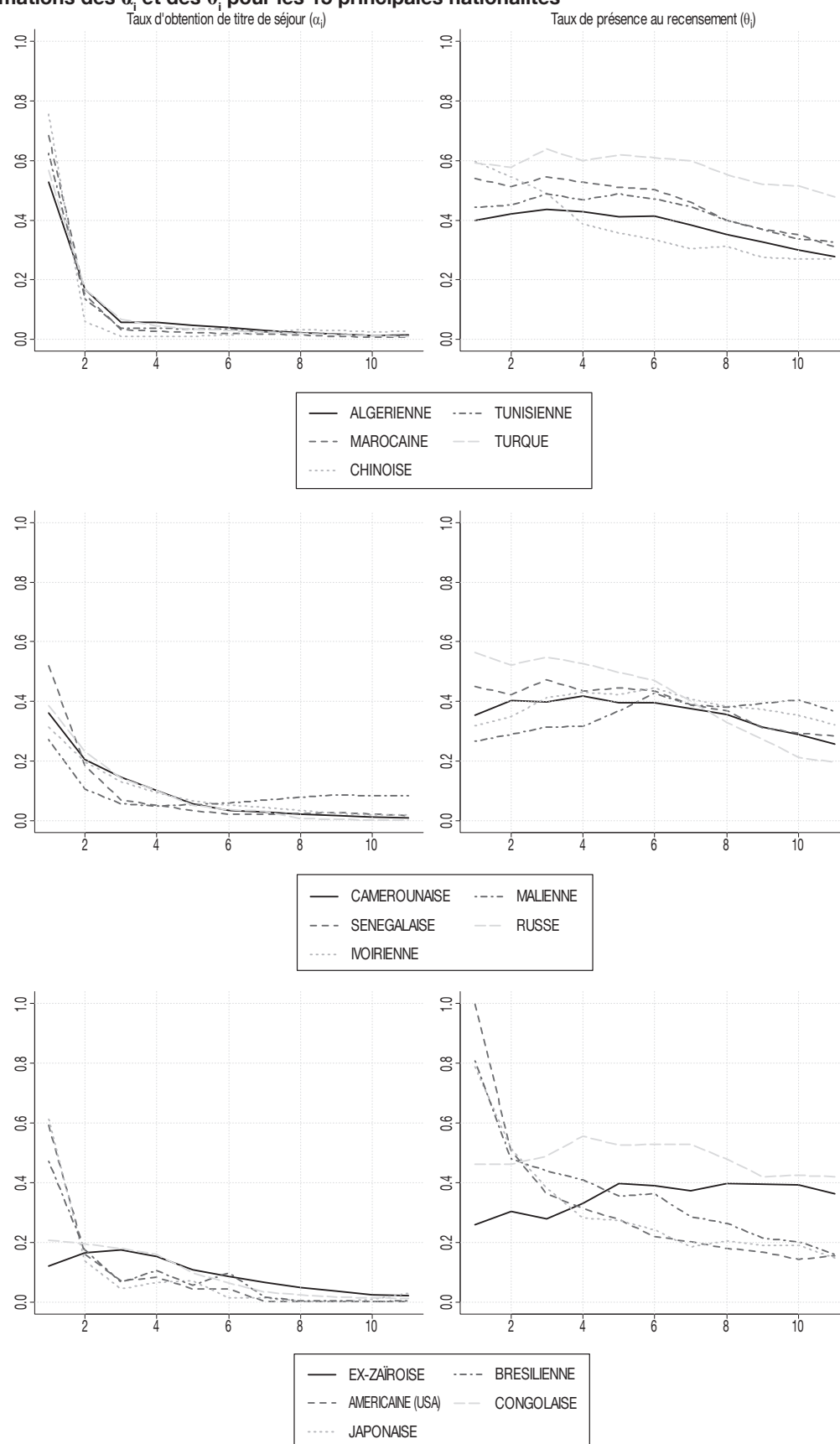
par an pour chaque pays (de moins de 1 % pour les États-Unis et le Mali à 4,4 % pour la Russie).

\* \*

\*

La base AGDREF constituée des titres de séjour délivrés aux migrants de nationalité de pays tiers ainsi que les enquêtes annuelles du *Recensement de la population* permettent de déterminer des flux de migrants vers la France métropolitaine. Pour pouvoir comparer ces deux sources, il faut se positionner sur le même champ, c'est-à-dire s'assurer qu'on mesure les mêmes flux avec l'une et avec l'autre des deux sources. Or le champ commun à ces deux sources, bien qu'il soit défini théoriquement, reste difficile à appréhender dans la pratique. Il est constitué des migrants de long terme (c'est-à-dire des migrants établissant, ou souhaitant établir, leur résidence habituelle pour une période d'au moins 12 mois sur le territoire français), âgés de 19 ans ou plus l'année de leur arrivée en France et dont la nationalité est celle d'un pays tiers. D'après la réglementation internationale, le recensement ne concerne que les personnes résidant depuis plus d'un an en France, ou qui y résideront plus d'un an. Le champ du recensement pour les migrants est donc celui des migrants de long terme. Néanmoins, il dépend de la prévision des habitants à rester plus d'un an ou non. Or cette prévision peut s'avérer fautive : une personne peut penser rester moins d'un an lors du recensement et finalement s'établir plus d'un an. De même, il est difficile de déterminer avec certitude les migrants de long terme dans la base AGDREF. En effet, cette base ne donne des informations que sur l'état administratif des migrants de pays tiers. Un migrant peut avoir un titre de séjour valide, mais ne pas séjourner en France. On a utilisé ici les renouvellements des titres de séjour ainsi que leur durée de validité pour savoir si un migrant est resté plus d'un an sur le territoire français ou non. Mais pour une partie des migrants, il est impossible de savoir avec certitude s'ils sont restés un an en France ou s'ils sont partis avant la fin de validité de leur titre en ayant séjourné moins d'un an. Certains choix ont donc été fait afin de déterminer les migrants de long terme. Même une fois ramené à un champ commun, on a décidé de ne pas comparer directement les flux mesurés par le recensement avec ceux constitués du nombre de titres de séjour délivrés chaque année. La raison en est qu'un titre de séjour n'est pas toujours délivré la même année que celle où le migrant entre sur le territoire. Il peut parfois s'écouler plusieurs années entre l'entrée en France et l'obtention du titre. Ainsi, les titres délivrés une

Figure VII  
Estimations des  $\alpha_i$  et des  $\theta_i$  pour les 15 principales nationalités



Lecture : ces courbes sont obtenues en estimant le modèle A séparément sur chaque sous-population avec  $\tau = 1$ .  
Champ : migrants de long terme, de nationalité de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine.  
Source : AGDREF, DSED et EAR, Insee, 2005-2012.

année concernent des migrants qui ne sont pas tous entrés la même année. De plus, si l'on veut avoir une source la plus exhaustive possible, il est nécessaire d'attendre plusieurs années, afin que tous les migrants aient eu le temps d'obtenir leur titre. Afin de surmonter ces difficultés et de rendre cohérentes les données issues des titres de séjour et celles issues des *EAR*, un modèle a été proposé, qui prend en compte le retard avec lequel certains migrants obtiennent leur titre. Ce modèle permet d'appréhender les deux sources comme étant la réalisation d'événements aléatoires, à savoir le fait d'être recensé et le fait d'obtenir un titre de séjour. Il fait intervenir des taux de présence au recensement ainsi que les vrais flux migratoires, qui ne sont pas directement observés. Les paramètres de ce modèle sont estimés de façon précise en utilisant l'inférence bayésienne, qui est un moyen efficace lorsque les paramètres sont liés entre eux, comme c'est le cas ici.

La principale limite du modèle est qu'il faut imposer le taux de migrants qui n'obtiendront

jamais de titre de séjour. Or ce taux est inconnu et ne peut pas, à notre connaissance, être estimé à partir d'autres sources. Par ailleurs, le modèle repose sur l'hypothèse que les taux  $\alpha_i$  et  $\theta_i$  sont les mêmes pour chaque génération entrée une année  $N$ . On peut penser qu'à long terme ou sous l'effet d'une réforme, ces taux évoluent de génération en génération, ce qui rendra les estimations du modèle moins bonnes. En particulier, l'introduction des VLS-TS à partir de 2009 modifie les taux  $\alpha_i$  d'obtention de titre et le changement de questionnaire des *EAR* à partir de 2011 peut avoir un impact sur les  $\theta_i$ .

Avec la méthode retenue ici, il est probable qu'on surestime les migrants de long terme dans la base AGDREF et qu'en conséquence on sous-estime les taux  $\theta_i$  de présence au recensement. D'autre part, près d'un quart des migrants des pays tiers ne répondent pas à la question des *EAR* portant sur l'année d'arrivée en France et 5 % ne répondent ni à l'année d'arrivée, ni à la question sur la résidence antérieure. Or, ce sont

#### Encadré 4

##### **LES TAUX DE NON-RÉPONSE À L'ANNÉE D'ENTRÉE SONT NÉGATIVEMENT CORRÉLÉS AUX TAUX DE PRÉSENCE AU RECENSEMENT ESTIMÉS.**

La non-réponse partielle à la question des *Enquêtes annuelles de recensement (EAR)* portant sur l'année d'entrée varie entre 15 % et un peu plus de 35 % selon la nationalité des migrants. On observe une corrélation négative entre ce taux de non-réponse et le taux de présence au recensement estimé pour la première année ( $\theta_1$ ). Plus le taux de non-réponse à la question sur l'année d'arrivée est important, plus le taux de présence estimé est faible (cf. figure). On pourrait alors penser qu'une partie des migrants n'est pas identifiée comme étant entrée l'année précédant le recensement du fait d'une mauvaise imputation de l'année d'arrivée. Le redressement de l'année d'arrivée prend en compte la question sur la résidence antérieure, ce qui a posteriori permet de limiter l'incertitude. Il reste que la variable indiquant le pays de résidence antérieure est elle-même une variable redressée dans le traitement des *EAR* : parmi ceux qui ont le pays de résidence antérieure redressé indiquant « France » et qui n'ont pas répondu à la question sur l'année d'arrivée, 29,4 % d'entre eux n'avaient en réalité pas répondu à la question sur le lieu de résidence antérieure (pour les *EAR* 2005 à 2010 ; ce chiffre est de 20,6 % pour les *EAR* 2011-2012). Parmi ceux qui n'ont pas répondu à l'année d'arrivée, 77 % ont néanmoins répondu à la question sur la résidence antérieure, ce qui permet de déterminer s'ils étaient en France ou non 5 ans (ou 1 an) avant la date du recensement.

Si on compare alors les  $\theta_1$  estimés pour différentes nationalités à la proportion de personnes n'ayant répondu ni à la question sur l'année d'arrivée ni à la question sur la résidence antérieure, on remarque encore une corrélation

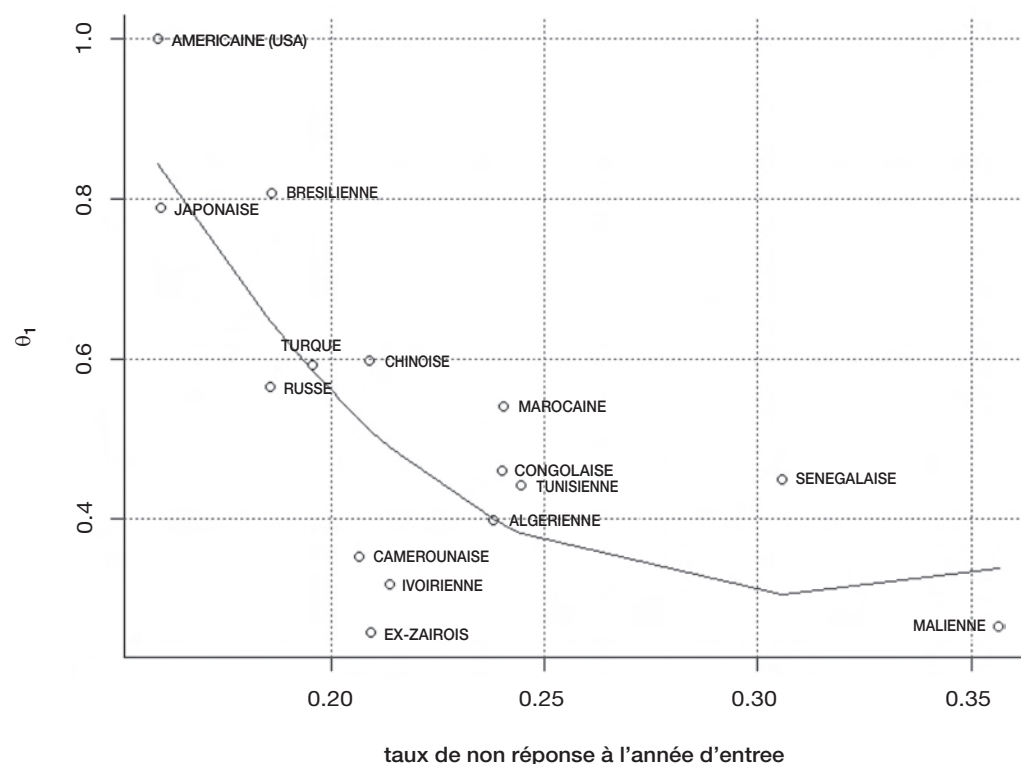
négative. Ce taux de non-réponse aux deux questions des *EAR* (année d'arrivée et résidence antérieure) est de l'ordre de 5 % sur les *EAR* 2005-2012. Actuellement, la résidence antérieure est redressée de façon automatique dans les traitements des *EAR* sans prendre en compte la nationalité des répondants, ce qui fait que plus de 99 % de ce qui n'ont répondu ni à l'année d'arrivée ni à la question sur la résidence antérieure sont redressés comme étant en France 5 ans (ou 1 an) avant la date du recensement. Par conséquent, ces migrants n'entrent pas en compte dans les flux d'entrées. La façon dont l'année d'arrivée de ces migrants est imputée a priori des répercussions sur les taux de présence au recensement estimés. Si par exemple on redresse la résidence antérieure de ces migrants, en prenant en plus en compte la nationalité, cela augmente au final le taux de présence au recensement de la première année  $\theta_1$  de 3 %. L'impact apparaît donc relativement faible. Néanmoins, si les migrants qui n'ont répondu ni à l'année d'arrivée ni à la question sur la résidence antérieure ne ressemblent pas à l'ensemble de la population, les redressements ainsi effectués peuvent être biaisés. Au final, une partie de la non-présence au recensement estimée pourrait en fait provenir d'une mauvaise identification des nouveaux migrants dans les *EAR* du fait d'une mauvaise imputation des années d'arrivée. Même dans le cas d'hypothèses plus extrêmes, comme par exemple considérer que tous ceux qui n'ont pas répondu à l'année d'arrivée et à la résidence antérieure sont entrés l'année précédant le recensement, les flux estimés ( $V_N$ ) n'augmentent pas beaucoup par rapport aux flux estimés dans cet article.



Encadré 4 (suite)

Figure

**Taux de présence au recensement estimé en fonction du taux de non réponse à la question portant sur l'année d'entrée en France (question des *EAR*), pour 15 pays tiers**



Lecture : la courbe a été obtenue par régression polynomiale.

Champ : migrants de long terme, de nationalité de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine.

Source : AGDREF, DSED et EAR, Insee, 2005-2012.

les réponses à ces deux questions qui permettent de déterminer les flux migratoires à partir des *EAR*. On a montré qu'il y a une corrélation négative entre le taux de présence estimé  $\theta_1$  et le taux de non-réponse à ces deux questions (cf. encadré 4), ce qui pourrait être le résultat d'une mauvaise imputation des années d'arrivée. Même si ces imputations ont été faites avec soin, en prenant en compte le maximum d'informations disponibles, il reste possible qu'elles donnent des résultats biaisés et conduisent à une sous-estimation des flux migratoires déterminés à partir des *EAR*. Les taux de présence au recensement sont donc entachés de ces deux incertitudes : celle liée aux migrants de long terme dans la base AGDREFF et celle liée à l'imputation des années d'arrivée dans les *EAR*.

Si l'on interprète les taux de présence au recensement donnés dans cet article sans prendre de précaution, on peut alors conclure que certains

migrants ne sont pas recensés et que les flux migratoires sont de ce fait sous-estimés. Pour autant, comme cela a été montré, une sous-estimation des flux de 40 %, telle qu'elle ressort dans cet article (soit environ 60 000 personnes) entraînerait une sous-estimation des stocks de l'ordre de 8 % seulement sur le champ des migrants de nationalité de pays tiers, âgés de 19 ans ou plus l'année de l'entrée en France (soit environ 150 000 personnes), ce qui constituerait une sous-estimation de l'ensemble de la population de l'ordre de 0,2 % (cf. encadré 5). L'erreur commise serait donc bien inférieure à la précision du *Recensement de la population*.

De nouvelles pistes sont à explorer pour mieux comprendre certains phénomènes. Il serait par exemple intéressant de savoir dans quelle mesure la détention d'un titre de séjour en cours de validité change le comportement envers le recensement en créant un modèle qui distingue



les individus qui ont un titre de séjour de ceux qui n'en ont pas au moment du recensement. Un autre aspect qui n'a pas été approfondi ici concerne les flux de sortie des migrants qui peuvent être appréhendés par les taux de

renouvellement des titres de séjour (pour les titres d'une durée de validité d'un an ou moins) d'une part, et par l'évolution dans les *EAR* des effectifs des populations entrées en fonction de leur temps passé en France d'autre part. □

#### Encadré 5

##### INFLUENCE DE L'ESTIMATION DES FLUX SUR L'ESTIMATION DES STOCKS

D'après la modélisation retenue dans cet article (imputation des années d'arrivées non renseignées dans les *EAR*, détermination des migrants de long terme selon les titres de séjour, modélisation probabiliste), les flux d'entrées de migrants de pays tiers nés à l'étranger, de 19 ans ou plus l'année de leur arrivée pourraient être sous-estimés par le *Recensement de la population*, de l'ordre de 60 000 personnes par an (ce chiffre est calculé en faisant la moyenne des  $V_N - E_{i,N}$  sur la période 2004-2011, les  $V_N$  étant estimés avec  $\tau=1$  – se reporter au tableau 5). Ceci n'implique en aucune façon que la population de migrants non-estimée par le recensement augmente de 60 000 personnes par an. En d'autres termes, l'incertitude sur les flux d'entrées n'a pas d'effet cumulatif sur les stocks. En effet, rien ne nous dit que les migrants qui ne sont pas présents au recensement lors de leur première année passée en France, ne seront pas présent au recensement par la suite. Il est même possible qu'une partie des migrants de long terme n'aient au départ tout simplement pas l'intention de rester en France plus d'un an : si tel est le cas, ils n'ont pas à se faire recenser la première année de leur arrivée en France.

La figure A illustre le fait que l'incertitude sur la mesure des flux d'entrées, bien qu'elle reste constante au cours du temps, ne se cumule pas au cours des années. Autrement dit, sous l'hypothèse d'un régime stationnaire sur les flux d'entrées non recensées, **la part de migrants non recensés n'augmente pas mais reste constante « en stock »**.

En France, aucun dispositif national ne permet de savoir lorsqu'une personne quitte le territoire de façon à ne plus résider habituellement en France. Une façon d'estimer les sorties à partir des titres de séjour est alors de déterminer les migrants qui, arrivés en fin de validité de leur dernier titre, ne le renouvellent pas. On peut alors penser que le migrant est sorti avant la fin de validité de son titre et après le dernier passage en préfecture enregistré (déterminé par les dates de dépôts des demandes). Dans la réalité, il est possible que le migrant reste illégalement sur le territoire après l'expiration de son titre, mais en l'absence d'information sur les refus de titre (y compris les refus de renouvellements de titre), on ne peut pas déterminer l'ampleur de ce phénomène.

En conséquence, on peut déterminer pour chaque migrant trois périodes : une première période pendant laquelle on peut être certain de la présence du migrant, suivie d'une période pour laquelle le migrant est potentiellement présent mais est peut-être déjà sorti du territoire, et enfin une période pour laquelle le migrant est

très probablement sorti du territoire. On appellera ici, « migrant certain » ceux qui sont dans la première période et « migrant incertain » ceux qui sont dans la deuxième.

Par exemple, si on compare, pour les migrants arrivés en 2004, le stock de migrants certains et incertains (source AGDREF) au stock de migrants recensés, on s'aperçoit que le recensement, tel qu'il est utilisé dans cet article, n'identifie pas bien certaines personnes lors des deux premières années après leur arrivée en France. À partir de la troisième année, la courbe du recensement se situe entre celle des migrants certains et des migrants incertains, ce qui est alors compatible avec le fait qu'à partir de 3 années passées en France, tous les migrants sont présents au recensement. Au vu de la figure B, l'hypothèse qu'au bout de 5 années passées en France tous les migrants sont bien présents au recensement, paraît raisonnable.

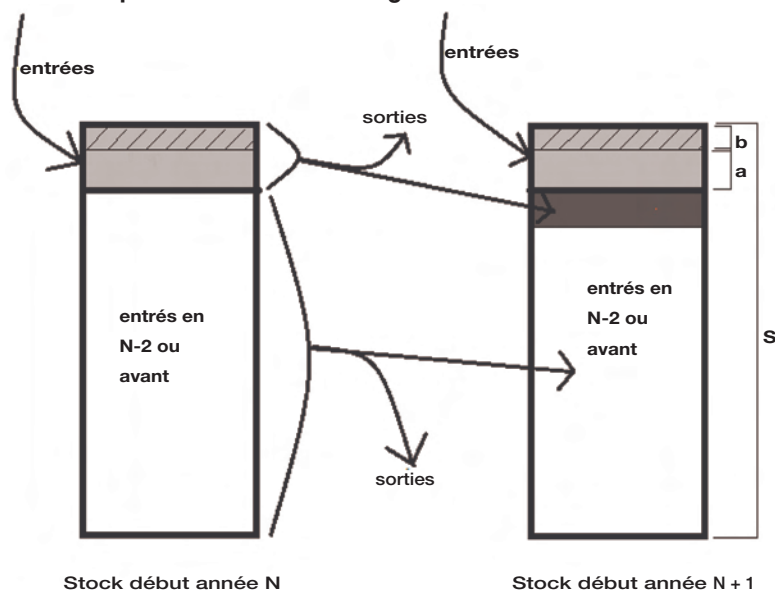
Quel est donc l'impact sur la mesure des stocks des migrants de pays tiers par le recensement si on retient cette hypothèse ? Dans ce cas il manque dans le stock d'une année  $N$ , déterminé par les *EAR*, une partie de ceux qui sont arrivés en  $N-1$ ,  $N-2$ ,  $N-3$  et  $N-4$ . En supposant que le taux de recensement augmente linéairement sur les 5 premières années, passant de 60 % (valeur estimée de 0) à 70 %, 80 %, 90 % puis 100 %, il manque dans le recensement effectué en janvier-février d'une année  $N$ , 40 % de ceux qui sont arrivés en  $N-1$ , 30 % de ceux qui sont arrivés en  $N-2$ , 20 % de ceux qui sont arrivés en  $N-3$  et 10 % de ceux qui sont arrivés en  $N-4$ . Or, les migrants arrivés en  $N-1$  représentent en moyenne (sur les *EAR* 2005-2012) 6,9 % du stock recensé, ceux arrivés en  $N-2$  5,7 %, ceux arrivés en  $N-3$  5,6 % et ceux arrivés en  $N-4$  5,5 %. Ceci implique que le recensement sous-estimerait alors le véritable stock des migrants de pays tiers de 8,3 %. (En désignant  $S$  le stock réel de migrants en janvier d'une année  $N$ ,  $a_i$  le nombre de migrants de ce stock entrés l'année  $N-i$  et  $b_i$  le nombre de migrants entrés l'année  $N-i$ , présents en  $N$ , mais non recensés (cf. figure A), on cherche alors à estimer le rapport  $(b_1 + b_2 + b_3 + b_4)/S$ . Il peut être déterminé en résolvant le système constitué des équations :  $a_i/(a_i + b_i) = \theta_i$  et  $a_i/(S - b_i) = \gamma_i$  où  $\gamma_i$  est la proportion moyenne, dans le recensement, des migrants arrivés l'année  $N-i$ .)

Or, dans le *Recensement de la population*, les personnes de nationalité de pays tiers, nées à l'étranger, entrées en France après l'âge de 19 ans, représentent 2,2 % de la population totale de France métropolitaine. Ainsi, une sous-estimation de 8,3 % d'entre elles revient à une sous-estimation de 0,2 % de l'ensemble de la population.



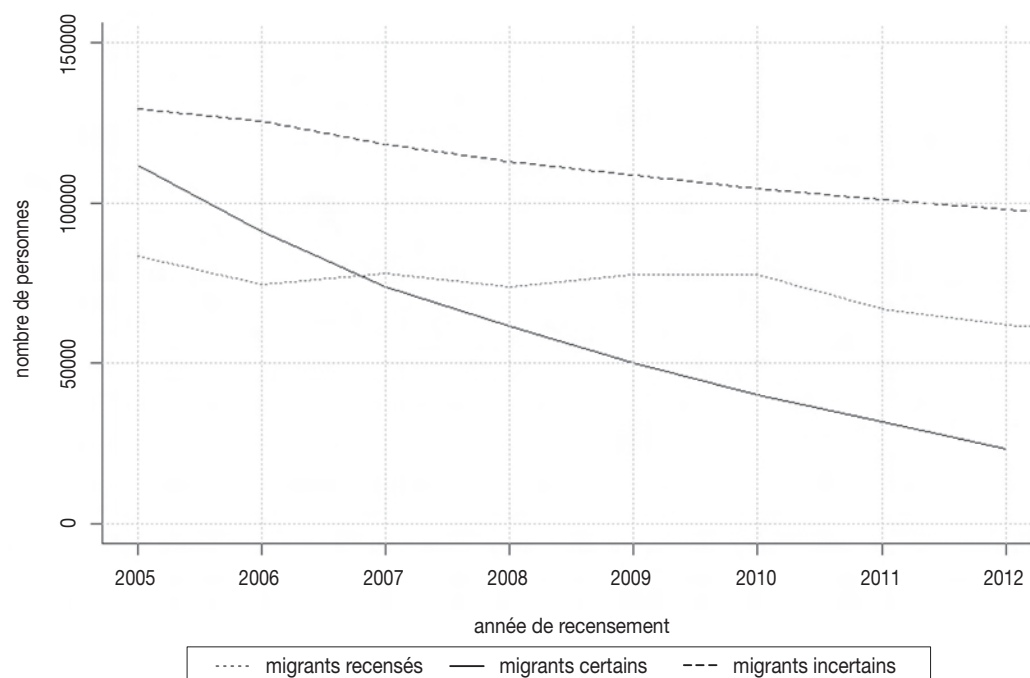
Encadré 5 (suite)

Figure A  
Illustration de la composition du stock de migrants en fonction de leur année d'arrivée



Lecture : les migrants arrivés l'année précédant le recensement, et qui constituent donc le flux d'entrée, sont représentés par le cadre du haut du stock : en gris hachuré ceux qui ne sont pas comptés dans le recensement, et en gris sans hachures ceux qui au contraire y sont comptés. Ce schéma illustre le fait que la part des migrants non recensés ne se cumule pas mais au contraire reste constante au cours du temps.

Figure B  
Évolution du stock de migrants de pays tiers entrés en France en 2004, selon qu'il est estimé selon le recensement, ou selon les titres de séjour  
Immigrants entrés en 2004 selon l'année du recensement



Lecture : les migrants « certains » une année N, sont ceux qui sont passés en préfecture après le 15 janvier de l'année N. Les migrants « incertains » sont ceux dont aucune trace de passage en préfecture n'a été enregistrée après le 15 janvier de l'année N, mais dont la fin de validité de leur titre de séjour a lieu après le 15 janvier de l'année N.  
Champ: migrants de long terme de nationalité de pays tiers, arrivés en France en 2004, âgés de 19 ou plus l'année de leur arrivée, France métropolitaine.  
Source : AGDREF, DSED et EAR 2005-2012, Insee.

## BIBLIOGRAPHIE

- Albert J.H. et Gupta A.K. (1983)**, « Estimation in contingency tables using prior information », *Journal of the Royal Statistical Society*, n° 1, pp. 60-69.
- Alkema L., Raftery A.E. et Gerland P. (2014)**, « Bayesian Population Projections for the United Nations », *Statistical Science*, n° 29 (1), pp. 58-68.
- Bègue M. (2007)**, « Qui sont les nouveaux bénéficiaires d'un titre de séjour en France ? », *Études et résultats*, Dres, n° 612.
- Bijak J. (2011)**, *Forecasting International Migration in Europe: A Bayesian View*, Springer Series on Demographic Methods and Population Analysis.
- Brutel C. (2014)**, « Les immigrés récemment arrivés en France », *Insee Première*, n° 1524.
- Brutel C. (2014)**, « Estimer les flux d'entrées sur le territoire à partir des enquêtes annuelles de recensement », *Document de travail*, n° F1403, Insee.
- Cadenel N. et Ménard S. (2010)**, « Peut-on compter le nombre de personnes qui entrent chaque année en France pour y vivre ? », *Regards croisés sur l'économie*, n° 8, pp. 212-217.
- Dares (2011)**, « Les nouveaux détenteurs d'un titre de séjour permanent en 2006 : une mobilisation importante des réseaux familiaux et personnels pour l'accès à l'emploi », *Dares Analyses*, n° 36.
- de Beer J., Raymer J., van der Erf R. et van Wissen L. (2010)**, « Overcoming the problems of inconsistent international migration data : a new method applied to flows in Europe », *European Journal of Population*, n° 26, pp. 459-481.
- Desrosières A. (2001)**, « Histoire de la raison statistique, le moment bayésien », *Courrier des statistiques*, n° 100.
- Mazuy M., Barbieri M. et d'Albis H. (2013)**, « L'évolution démographique récente en France : la fécondité est stable », *Population*, n° 68, vol. 3, pp. 385-432.
- Ménard S. et Papon S. (2011)**, « Le devenir des étudiants étrangers en France », *Infos migrations*, n° 29.
- Nations Unies (1994)**, « Rapport de la conférence internationale sur la population et le développement ».
- Nations Unies (2013)**, « The number of international migrants worldwide reaches 232 million », *Population Facts*, n° 2013/2, Department of Economic and Social Affairs, Population Division.
- ONU (1999)**, « Recommandations en matière de statistique des migrations internationales », *Études statistiques*, n° 58, série M, Rev. 1, Départements des affaires économiques et sociales, Division de la statistique.
- Poulain M. et Herm A. (2002)**, « Les flux migratoires internationaux en Europe », *Futuribles*, n° 279, pp. 5-27.
- Raymer J. et Abel G. (2008)**, « The MIMOSA model for estimating international migration flows in the European Union », *Document de travail*, Genève, UNECE / Eurostat.
- Raymer J., Abel G. et Smith P.W.F (2007)**, « Combining census and registration data to estimate detailed elderly migration flows in England and Wales », *Journal of Royal Statistical Society, Series A.*, n° 170, pp. 891-908.
- Raymer J., Forster J.J., Smith P.W.F., Bijak J. et Wisniowski A. (2011)**, « Integrated modelling of european migration : background, specification and results », Imem Workshop, Chilworth.
- Sardon J.P. (2006)**, « Évolution démographique récente des pays développés », *Population*, n° 61, pp. 225-300.
- Secrétariat général à l'immigration et à l'intégration (2013)**, *La délivrance des visas aux étrangers*, 2013. <http://www.immigration.interieur.gouv.fr/content/download/39276/300376/file/Visas.pdf>
- Thierry X. (2001)**, « Les entrées d'étrangers en France de 1994 à 1999 », *Population*, n° 56, vol. 3, pp. 423-450.
- Union européenne (2008)**, règlement (CE) no 763/2008 du parlement européen et du conseil du 9 juillet 2008 concernant les recensements de la population et du logement., 2008.
- Willekens F. (1994)**, « Monitoring international migration flows in Europe, towards a statistical data base combining data from different sources », *European Journal of Population*, n° 10, pp. 1-42.

## ROBUSTESSE DU MODÈLE

Dans cette partie on a simulé des données selon le modèle A avec les paramètres suivants :  $\alpha_1=0,5$ ,  $\alpha_2=0,1$ ,  $\alpha_3=0,05$ ,  $\alpha_4=0,05$ ,  $\alpha_5=0,03$ ,  $\alpha_6=0,02$ ,  $\alpha_7=0,02$ ,  $\alpha_8=0,01$ ,  $\alpha_9=0,01$ ,  $\alpha_{10}=0,005$ ,  $\alpha_{11}=0,005$  et  $\alpha_j=0$  pour  $j \geq 12$  ;  $\theta_1=0,6$ ,  $\theta_2=0,55$ ,  $\theta_3=0,5$ ,  $\theta_4=0,45$ ,  $\theta_5=0,4$ ,  $\theta_6=0,35$ ,  $\theta_7=0,3$ ,  $\theta_8=0,3$ ,  $\theta_9=0,25$ ,  $\theta_{10}=0,25$ ,  $\theta_{11}=0,22$ . On a tiré la population totale entrante une année  $N$  selon une loi normale :  $V_N \sim N(V^*, 2000)$ .

On a ensuite estimé ce modèle, sans fixer  $\sum_{i=1} \alpha_i$ , pour différentes valeurs de  $V^*$  et différentes valeurs de la borne supérieure  $V_{sup}$  sur l'a priori des  $V_N$  afin de voir quelle est son influence sur les estimations. Par la suite, on note  $V_{est}$  la moyenne des  $V_N$  estimés.

On voit d'après la figure A que les estimations ne convergent pas vers les vraies valeurs du modèle. En effet,  $V_{est} \approx (V_{inf} + V_{sup})/2$ , ce qui indique qu'il n'y a pas assez d'informations disponibles dans le modèle pour retrouver la bonne valeur de  $V^*$ . Ceci n'est pas dû au manque d'observations, car même si on augmente

considérablement le nombre d'observations, on retrouve le même type de figure.

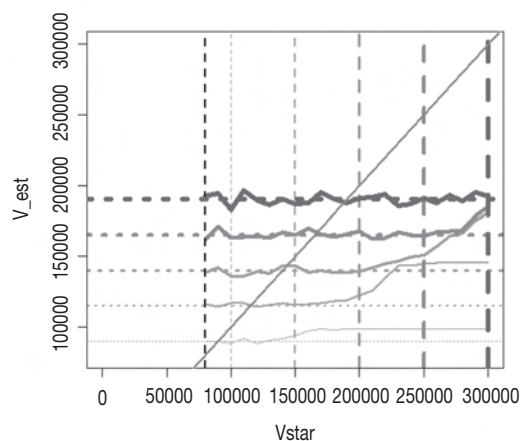
On refait les mêmes simulations et estimations, mais cette fois-ci en connaissant à l'avance  $\tau = \sum_{i=1}^{11} \alpha_i$ . Dans

le modèle simulé,  $\tau = 0,8$ . Si on rajoute cette contrainte au modèle, on trouve alors les estimations de la figure B qui montrent que lorsque la vraie valeur  $V^*$  se trouve entre les bornes  $V_{inf}$  et  $V_{sup}$  de l'a priori, les estimations convergent vers les vrais paramètres. Si  $V^*$  est supérieur à  $V_{sup}$  (cas d'une borne supérieure de l'a priori trop faible) alors les estimations convergent vers  $V_{sup}$ .

La figure C montre comment varient les estimations en fonction de la valeur de  $\tau$  imposée au modèle (la vraie valeur de  $\tau$  reste ici 0.8). Les estimations sont alors égales aux vrais paramètres lorsqu'on impose le bon  $\tau$  (lorsque  $1 - \tau = 0,2$ ). Sinon, les estimations des  $\alpha_i$  et  $\theta_i$  évoluent de façon linéaire avec  $\tau$ .

Figure A

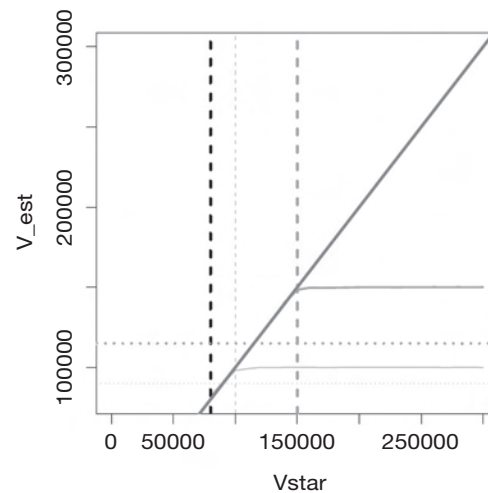
**Moyenne des estimations des  $V_N$  sur des données simulées selon le modèle A, en fonction de la vraie valeur moyenne des flux de migrants  $V^*$  et cela pour différentes borne supérieures de l'a priori sur les  $V_N$**



Lecture : du trait le plus fin au trait le plus épais, la borne supérieure de l'a priori vaut successivement 100 000, 150 000, 200 000, 250 000 et 300 000. La borne inférieure reste fixée à 80 000 dans tous les cas. On a rajouté la première bissectrice permettant de comparer l'écart entre l'estimation  $V_{est}$  et la vraie valeur  $V^*$ . Les traits horizontaux correspondent à  $(V_{inf} + V_{sup})/2$  et les traits verticaux à  $V_{sup}$ . Données simulées.

Figure B

**Moyenne des estimations des  $V_N$  sur des données simulées selon le modèle A, en fonction de la vraie valeur moyenne des flux de migrants ( $V^*$ ) et cela pour différentes bornes supérieures de l'a priori sur les  $V_N$**

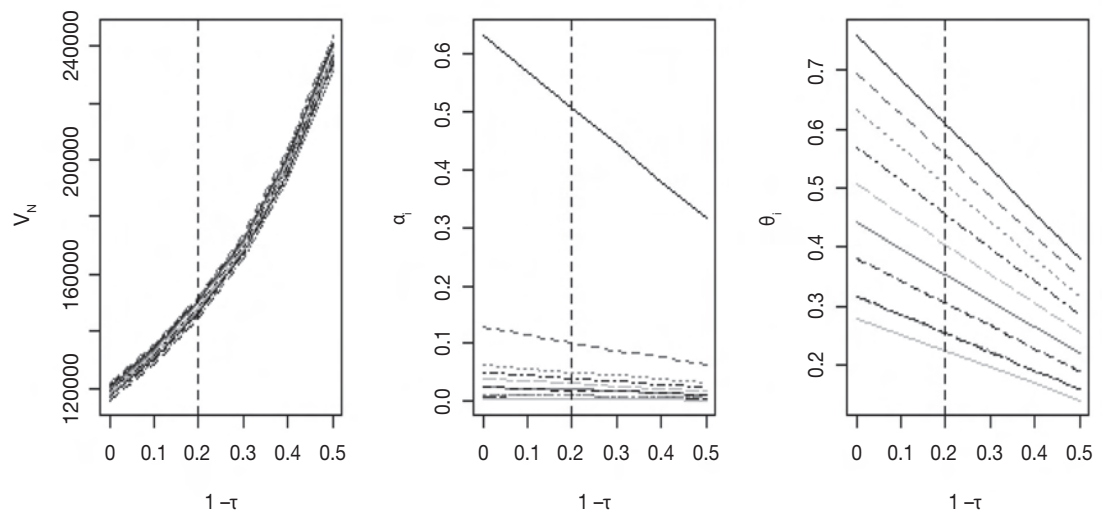


Lecture : la borne supérieure de l'a priori vaut 100 000 (trait fin) ou 150 000 (trait épais). La borne inférieure est fixée à 80 000 dans tous les cas. On a ajouté la première bissectrice permettant de comparer l'écart entre l'estimation  $V_{est}$  et la vraie valeur  $V^*$ . On contraint de plus dans les estimations à avoir  $\tau=0,8$ . On remarque alors que lorsque  $V^*$  est inférieur à la borne supérieure,  $V_{est}=V^*$  ( $V_{est}$  est sur la première bissectrice).

Données simulées.

Figure C

**Sensibilité des estimations des paramètres du modèle A en fonction de la proportion supposée de migrants entrant une année donnée et qui n'auront jamais de titre ( $1 - \tau$ )**



Lecture : les trois graphiques représentent les trois séries de paramètres (les  $V_N$ , les  $\alpha_i$  et les  $\theta_i$ ) estimés par le modèle en fonction de la valeur de  $\tau$  fixée. Ici les données ont été simulées avec un vraie  $\tau^*$  qui vaut 0,8 (donc  $1 - \tau^* = 0,2$ ). On remarque que si on fixe un  $\tau$  différent du vrai  $\tau^*$  on estime mal les vrais paramètres : les  $V_N$  estimés augmentent avec  $1 - \tau$  tandis que les  $\alpha_i$  et  $\theta_i$  estimés diminuent avec  $1 - \tau$ .

Données simulées.



TABLES DES  $E_{i,N}$  ET DES  $A_{i,N}$  UTILISÉES

$A_{i,N}$  : nombre des migrants de nationalité de pays tiers, nés à l'étranger, étant arrivés en France métropolitaine l'année N à l'âge de 19 ans ou plus pour y résider au moins 12 mois, et dont la validité de leur premier titre de séjour commence l'année N+i-1.

$E_{i,N}$  : nombre des migrants de nationalité de pays tiers, nés à l'étranger, étant arrivés en France métropolitaine l'année N à l'âge de 19 ans ou plus pour y résider au moins 12 mois, et étant recensés l'année N+i.

Tableau A  
Table des  $A_{i,N}$

N																		
	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	78 280	75 059	72 615	73 985	83 369	86 658	87 977	86 229
2	NC	NC	NC	NC	NC	NC	NC	NC	NC	24 916	23 844	21 055	17 751	17 639	19 042	13 544	12 840	NC
3	NC	NC	NC	NC	NC	NC	NC	NC	11 817	10 421	8 307	7 326	6 214	6 358	6 910	7 750	NC	NC
4	NC	NC	NC	NC	NC	NC	NC	11 455	8 629	6 792	5 225	5 502	4 824	4 462	4 544	NC	NC	NC
5	NC	NC	NC	NC	NC	NC	6 262	8 382	6 973	4 510	4 424	4 050	3 151	2 800	NC	NC	NC	NC
6	NC	NC	NC	NC	NC	3 812	4 829	7 799	4 746	4 106	3 821	3 182	2 328	NC	NC	NC	NC	NC
7	NC	NC	NC	NC	2 189	3 033	4 903	5 190	4 203	3 731	3 121	2 525	NC	NC	NC	NC	NC	NC
8	NC	NC	NC	1 158	1 622	3 147	2 890	5 160	3 924	2 964	2 378	NC	NC	NC	NC	NC	NC	NC
9	NC	NC	624	846	1 760	1 760	2 930	4 735	3 066	2 320	NC	NC	NC	NC	NC	NC	NC	NC
10	NC	532	489	880	958	1 753	2 716	3 835	2 460	NC	NC	NC	NC	NC	NC	NC	NC	NC
11	779	668	794	667	1 176	2 085	2 812	3 812	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
12	692	616	514	689	1 269	2 100	2 785	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC

Lecture : « NC » indique que le chiffre n'est pas connu.

Champ : migrants de long terme, de nationalité de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine.

Source : AGDREF, DSED.

Tableau B  
Table des E<sub>i,N</sub>

		N																		
		1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
i	1	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	83 377	83 639	79 680	82 972	82 667	92 782	98 261	93 393	97 265
	2	NC	NC	NC	NC	NC	NC	NC	NC	NC	78 300	74 459	72 846	70 587	70 390	80 228	86 822	88 005	91 250	NC
	3	NC	NC	NC	NC	NC	NC	NC	NC	88 646	80 661	77 976	72 592	68 675	70 288	79 122	82 853	90 354	NC	NC
	4	NC	NC	NC	NC	NC	NC	NC	92 184	79 334	77 799	73 621	72 034	69 098	68 384	76 329	80 073	NC	NC	NC
	5	NC	NC	NC	NC	NC	NC	NC	80 821	87 359	81 949	77 606	73 844	65 912	65 530	75 997	NC	NC	NC	NC
	6	NC	NC	NC	NC	NC	NC	61 364	72 935	91 368	81 855	75 765	77 542	68 524	63 431	62 978	NC	NC	NC	NC
	7	NC	NC	NC	NC	NC	39 861	54 067	69 280	85 873	74 942	72 752	66 721	67 829	60 356	NC	NC	NC	NC	NC
	8	NC	NC	NC	NC	25 260	35 604	49 815	67 145	80 839	73 417	65 674	61 872	62 635	NC	NC	NC	NC	NC	NC
	9	NC	NC	NC	19 154	20 986	34 396	46 801	63 586	78 752	65 870	64 015	59 279	NC	NC	NC	NC	NC	NC	NC
	10	NC	NC	18 415	18 236	20 654	34 288	46 672	65 551	74 783	65 769	62 358	NC	NC	NC	NC	NC	NC	NC	NC
	11	NC	19 717	19 466	17 467	21 393	32 656	42 304	61 592	71 054	61 793	NC	NC	NC	NC	NC	NC	NC	NC	NC
	12	694 229	649 771	657 832	643 592	621 327	639 076	633 766	675 815	743 181	0	NC	NC	NC	NC	NC	NC	NC	NC	NC

Lecture : « NC » indique que le chiffre n'est pas connu.

Champ : migrants de long terme, de nationalité de pays tiers, nés à l'étranger, de 19 ans ou plus l'année de leur arrivée, France métropolitaine.

Source : EAR, Insee, 2005-2012.

